# Dealing with setbacks

*Exploring academic buoyancy in Latin students who are cognitively gifted*

O'Brien, C.

**Publication date**
2023

**Document Version**
Final published version

# DEALING WITH SETBACKS

CHELSEA O'BRIEN

*Exploring Academic Buoyancy in Latin Students who are Cognitively Gifted*

CHELSEA O'BRIEN – DEALING WITH SETBACKS

*Exploring Academic Buoyancy in Latin Students who are Cognitively Gifted*

UNIVERSITEIT VAN AMSTERDAM

DEALING WITH SETBACKS

EXPLORING ACADEMIC BUOYANCY IN LATIN STUDENTS
WHO ARE COGNITIVELY GIFTED

UNIVERSITY OF AMSTERDAM
Research Institute of Child Development
and Education

Dealing with Setbacks.
Exploring Academic Buoyancy in Latin Students who are Cognitively Gifted
ISBN: 978-94-6473-229-0
Cover design: Maria Heikens, Groningen
Printing: Ipskamp Printing


This publication (with project number 023.006.037) is part of the research programme Doctoral Grant for Teachers which is financed by the Dutch Research Council (NWO).

DEALING WITH SETBACKS
EXPLORING ACADEMIC BUOYANCY IN LATIN STUDENTS
WHO ARE COGNITIVELY GIFTED

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 18 oktober, 11.00 uur

door

Chelsea O’Brien
geboren te Barrow-in-Furness

Promotiecommissie

| | |
|---|---|
| Promotor: | prof. dr. G. C. W. Rijlaarsdam<br>Universiteit van Amsterdam |
| Co-promotor: | dr. S. M. Adema<br>Universiteit Leiden |
| Overige leden: | prof. dr. M. L. L. Volman<br>Universiteit van Amsterdam |
| | prof. dr. S. J. Andringa<br>Universiteit van Amsterdam |
| | prof. dr. E. H. Kroesbergen<br>Radboud Universiteit |
| | dr. C. K. K. Zanolie<br>Universiteit Leiden |
| | dr. L. W. van Gils<br>Universiteit van Amsterdam |

Faculteit der Maatschappij- en Gedragswetenschappen

# TABLE OF CONTENTS

# INTRODUCTION

Since I started teaching in 2011, I have been intensely involved with teaching and mentoring gifted students. Within the Netherlands, our secondary school has uniquely offered full-time gifted education since 2011 to most of our gifted students in the first three years and a specialized gifted mentoring program in the final three years. One of the foundations of our extensive gifted program is to not only focus on the cognitive development of our gifted students, but more importantly to look at their identity development as a whole, which includes for example metacognitive and affective strategies. To do so, we try not to approach behaviour exhibited by the students as attentive or disruptive, but to think of behaviour in terms of helpful or unhelpful strategies. This approach has resulted in mostly success stories over the last 12 years.

However, despite many success stories, each year we are still confronted with students with a less positive story. The students underachieve because we are not able to coach them into really engaging with learning. We do not seem to succeed in helping them adopt more helpful strategies.

In some particularly complex cases, the students even ended up dropping out of secondary school. These examples are few and far between, but they do occur. Moreover, each year we are approached by other schools, psychologists and parents of children who are not our students for advice on helping gifted students who are chronically underachieving or have dropped out of school. We frequently (attempt to) reintegrate gifted dropouts in secondary education. Such reintegration programs come at great costs, both financially and timewise for all parties concerned and, in our experience, are often not as successful as one would hope. Therefore, avoiding gifted students (chronically) underachieving and in extension dropping out in the first place seems pivotal for schools and their teachers.

On a smaller scale, during my own Latin and Greek lessons, particularly when we are completing a translation exercise, underachievement and disengagement from the task are frequently evident among the students, regardless of how proficient they are. When I ask them what is wrong, their first answer is usually that it is 'boring'. However, over the years I have discovered through further questioning that 'boring' more often than not is a codeword for anything that is uncomfortable. For example, the student might actually find the task difficult, or is scared of making a mistake, but instead of saying that, they say they are bored. Calling a task boring is easy. It is a statement, and nothing can be done to change it. At that moment, the learning stops. The range of strategies that students employ to get out of the discomfort and, in extension, learning fascinates me. I consider it my task as a teacher to figure out what is causing the student to (unconsciously) avoid learning and turn the discomfort into a learning experience.

As a teacher, I have often reflected on why some students successfully overcome their discomfort and learn, whereas others steadfastly do not. Questions frequently arose such as: 'why do some of my students not progress in translating and why do they keep using the same unhelpful strategies?' or 'why did student X decide to give up on Latin after one bad grade, whereas others keep on struggling for six years?'. Research in (Dutch) classrooms related to such questions was non-existent to my knowledge and, generally, educational based research with gifted participants remains scarce. Despite classical languages often being offered to high potentials to cognitively challenge them, we are actually still very much in the dark regarding why some students grasp the opportunity for being challenged and others avoid this.

The questions I ask myself as a teacher of classical languages are not the only questions pertaining to why some gifted students overcome discomfort and thrive at learning whereas others do not. From a mentor perspective, related but more general questions frequently come to mind such as 'how we can help our gifted students to thrive instead of underachieving?' or 'how can we avoid gifted students from dropping out of secondary school?'. As mentors of gifted students, we are still acting in the dark and basing our actions more on intuition than on evidence-based research. This is problematic, as each year in the Netherlands, too many gifted students either underachieve or drop out of school, with an estimated 1500-2500 gifted students receiving no education as they have dropped out (Zonnenberg, 2022). The need for answers provided by research, thus, seemed paramount.

This made me shift my perspective to that of a researcher and this dissertation came into being. The aim was to establish how a too difficult translation task (i.e., a short-term setback) affected cognitively gifted students' translation processes during and after the setback. We assumed that, if we could gain insights into how discomfort affects learning in different students, we should be able to design interventions to strengthen students' learning processes. Originally, this dissertation was intended as an intervention study. However, the collected data proved to be of such richness and complexity that a follow-up intervention study was a bridge too far and now lays outside the scope of the current dissertation. The study in its current form provides new insights into the employed strategies employed by our gifted participants and how they were affected by an academic setback. Therefore, we are hopeful that it provides a small piece of the puzzle as to why some gifted students thrive and others do not.

This piece of the puzzle is not only desirable from an education perspective, but also from a political one. In 2021 the *College voor Rechten van de Mens* (CRM, the Dutch College for Human Rights) strongly advised the Dutch government to take measures to ensure that gifted students' right to tailored education is met (Article 28).[1] Moreover, the coalition agreement drawn up by the government at the time of writing, Rutte IV, aims to provide all students, including gifted students, tailored education. The government is also aiming to significantly decrease the number of gifted students not receiving education. Furthermore, the then Minister of Education (Dennis Wiersma) drastically increased the budget for gifted education, raising from 14 million euros in 2022 to 24,5 million euros in 2023. This is to be further increased to 28 million euros in 2024. For gifted students, these are all positive developments. However, without the accompanying research, pertaining to the underlying factors that lead to some gifted students underachieving and dropping out from secondary education, whereas other thrive, educators cannot employ the new funding and facilities to create and implement successful interventions.

Even with the general need for answers regarding why some gifted students academically thrive and others do not, this dissertation was mainly inspired by the success stories of many of my gifted students on the one hand. On the other hand, it was driven by the often heart wrenching stories of those who were less successful. Gifted students and their stories, future and past, thus form the heart of this dissertation. On a larger scale, more insights into why some thrive and others do not would be more widely benefit the world of education. This is because, underachievement and dropping-out are not just problems faced by gifted students in the Netherlands, nor are they new problems (e.g. Gross, 1993; Pfeiffer & Stocking, 2000; Reis & Renzulli, 2004; Barbier, Donche, Verschueren,

---

[1] *"Hoogbegaafde kinderen hebben net als ieder ander kind <u>recht op onderwijs dat past bij hun onderwijsbehoeftes en leerpotentieel</u>. Dit recht moet altijd centraal staan in het onderwijsbeleid. Hiervoor moet de regering de nodige maatregelen treffen die de ondersteuning voor hoogbegaafde kinderen borgen. Denk bijvoorbeeld aan maatregelen die ervoor zorgen dat leraren meer tijd hebben voor ondersteuning en meer specifieke expertise kunnen vergaren over de leerbehoeftes van deze leerlingengroep. In dat verband vindt het College het ook belangrijk dat er geen extra kosten of andere (financiële) drempels worden opgeworpen voor de ondersteuning van deze kinderen." Quoted from a letter from the CRM to the caretaker minister of Education, Arie Slob, dated July 2nd, 2021. Retrieved from: https://publicaties.mensenrechten.nl/publicatie/5d15e9a4-320e-459e-9a9d-91bb83c88993*

2019). Therefore, this study is also situated in the larger perspective of gifted education research.

The gifted students included in this study were selected by means of a non-Latin yet linguistic task including a setback. Our 16 participants were selected because they were either performed significantly worse after this setback or significantly better. These participants then translated nine Latin sentences, while thinking aloud. Three of these sentences were, however, untranslatable, functioning as a setback. The task was immediatly followed by a retrospective interview in which the participants reflected on how they had translated the task.

The research questions and hypotheses underlying this experiment connect several concepts and ideas related to giftedness and setbacks in educational settings. Before presenting our research questions, we first introduce these concepts and explain our choice for Latin as the domain of this study.

## 1 ACADEMIC BUOYANCY

Learning is accompanied by discomfort and mistakes. It is helpful when students are able to bounce back and continue their school work after a setback, such as a particularly difficult exercise or another uncomfortable experience. Underachievement and dropping out seem related to students' abilities to bounce back from setbacks. When referring to the ability to bounce back from daily academic setbacks, this is known as academic buoyancy (Martin & Marsh, 2008; Martin & Marsh, 2009; Martin, 2013). Examples of such setbacks include receiving a disappointing grade, incidental underachieving or having to perform a difficult task, such as translating Latin. Students who are non-buoyant are known to experience more negative emotions than those who are academically buoyant (Fried & Chapman, 2012) and are more likely to avoid challenges and to develop a fear of failure (Meneghel, Martínez, Salanova & de Witte, 2019). In the long term, a lack of academic buoyancy can develop into a lack of academic resilience. Academic resilience refers to the ability to deal with long-term school adversities that pose major threats (Martin & Marsh, 2009). A lack of academic resilience could in turn lead to chronic underachievement and dropping out (Martin & Marsh, 2009; Martin, 2013, Agasisti, Avvisati, Borgonovi, & Longobardi, 2018).

The description of a 'setback' is reminiscent of what we previously described as 'discomfort' and 'bouncing back' was reminiscent of the 'thriving' seen in some of my gifted students. Moreover, as a teacher I regularly see my students experience negative emotions or try to avoid problems. It seemed then that academic buoyancy might to some extent be at play in answering why some gifted students thrive, whereas others do not.

Luckily, the extent of academic buoyancy in students can be improved (Martin, Colmar, Davey & Marsh, 2010). However, for Dutch teachers to create subject specific interventions to improve academic buoyancy, they must be able to recognise learning strategies related to low and high academic buoyancy. As we further discuss in Chapter 1, thus far, academic buoyancy has been studied by a somewhat limited group of researchers, mainly in Australia and the UK (e.g. Martin & Marsh, 2008; Martin & Marsh, 2009; Martin, 2013). Furthermore, these researchers have always relied on collecting their data via questionnaires, usually via self-report (eg. Malmberg, Hall & Martin, 2013; Kim & Han, 2014; Comerford, Batteson & Tormey, 2015; Jahedizadeh, Ghonsooly & Ghanizadeh, 2019) and it has been demonstrated that teachers are unable to identify behaviour related to academic buoyancy (Verrier, Johnson & Reidy, 2018). Therefore, we chose academic buoyancy as the central topic of this dissertation. We focus on how a lack of academic buoyancy in gifted students is reflected in their task processes.

## 2 EXPERIENCING SETBACKS THROUGH CHALLENGING TASKS: BENEFITS AND PREREQUISITES

Academic challenges are essential to learning. In this section we explore how gifted students might benefit from being challenged and what is necessary for them to benefit from challenges. The necessity and general benefits of being challenged have regularly been discussed in relation to gifted students (e.g. Eyre, 1997; Wallace 2000). In the coming paragraphs, we first discuss the benefits of challenges related to motivation. We then discuss the benefits to students' development of metacognitive skills.

Research has shown that being academically challenged increases motivation in gifted students (e.g. Lens & Rand, 2000). This is relevant as motivation contributes to the success of gifted students and, like academic buoyancy, seems to be related to underachievement (e.g. Albaili, 2003; Baslanti & McCoach, 2006; Matthews & McBee, 2007). In maths, gifted students have been seen to be more motivated and committed to learning when they are given challenging problems to solve (Diezmann & Watters, 2002). Furthermore, despite the lack of a general definition of giftedness (for more on this, see Chapter 2), motivation is a common factor in many giftedness models (e.g. Renzulli, 1978; Gagné, 2000; Subotnik, Olszewski-Kubilius & Worrell, 2011; Ziegler & Phillipson, 2012). For example, in the 2005 modification of Renzulli's giftedness model, the term motivation was specified as 'task commitment'. 'Task commitment' refers to a specific type of motivation that is necessary to complete challenging tasks. Task commitment is also sometimes referred to as ‘perseverance’ (Renzulli & Reis, 2018). Thus, motivating and challenging gifted students seem linked and necessary to training academic buoyancy and, in the long run, promoting their academic success.

Challenging gifted students may also have beneficial effects on their metacognitive skills. When tasks are too easy, gifted students can solve them without using metacognitive skills (Veenman, 2011). Challenging tasks, on the other hand, invite students to monitor and reflect upon their learning processes. A student who has developed metacognitive skills due to previous encounters with difficult tasks, may approach a challenging task by planning, defining which problems need to be solved, using different strategies and monitoring their progress before reflecting upon their work (Beyer, 2000). In contrast, a student who has not frequently encountered challenging tasks is unlikely to employ various strategies to solve the problem.

Dutch gifted students are known to frequently have metacognitive deficiencies due to a lack of frequent academic challenges (Veenman, Kerseboom & Imthorn, 2000). This is problematic, and relevant to our study, as a lack of metacognitive skills seems to be one of the underlying factors of underachievement (e.g., Carr, Borkowski & Maxell, 1991).

Challenging (gifted) students is therefore beneficial to their learning processes. However, we assumed that only challenging them is not sufficient, because gifted students who are not academically buoyant are likely to not persevere through the challenges and thus benefit from them. Therefore, to help gifted students benefit from difficult tasks, we need to better understand how their learning processes are affected by difficult tasks and how this relates to their extent of academic buoyancy.

## 3 CHOICE OF LATIN AS THE DOMAIN

For this study, we needed a task-based measure of academic buoyancy for reasons explained below. To do so, we developed a task based on translating Latin sentences into Dutch (see Chapter 1). As a teacher of classical languages, using a Latin translation task seemed a logical step for that reason alone. In addition, we had three reasons to suppose that a such a task would be particularly relevant and fitting for studying academic buoyancy in gifted students.

First, research on gifted education or metacognition tends to focus on the STEM subjects and not on languages. For maths, there are already studies on how teachers can support their students when having to perform a challenging task (e.g. Sullivan, et al., 2015). Linguistic subjects are represented less in these research fields, despite opportunities for problem solving tasks in linguistic domains. An example of linguistic problem solving that has received some attention is anagram solving (Lucas, Gratch, Cheng & Marsella, 2015). However, anagram solving is not part of school curricula. Thus, an anagram-based setback task would not be representative of a setback that students encounter in an educational setting. Translating Latin texts on the other hand is a task that students regularly perform at school.

A second reason for Latin translation being a suitable domain for studying academic buoyancy, was because translation is a demanding and complex activity. Generally, translation simultaneously relies on various cognitive and metacognitive activities (Göpferich, Jakobsen & Mees, 2009; Luger, 2020). Specifically translating Latin makes further demands on cognition: the text structures, reasoning paths and content are unfamiliar to the modern learner (Boyd, 2018; Florian, 2015). This means that students are frequently relying on their problem-solving skills to navigate the complexity of the task, again making this an interesting activity for the current research.

Finally, the texts used in the Latin secondary school curriculum (upper levels) themselves provide a reason for a Latin translation task to be fitting for the current research. One of the aims of the Latin curriculum is that students are able to translate authentic Latin literature. Often, the source texts are heterogeneous regarding their complexity. A simple sentence might be followed by a particularly complex sentence containing many grammatical features, only to be followed again by a simple sentence. Moreover, the texts the students are expected to translate are so called 'frustration level texts', always containing unfamiliar grammatical and vocabulary elements (Boyd, 2018). This heterogeneous and difficult nature of the Latin source texts suits the idea of mimicking a typical school situation where a student might benefit from picking themselves up after a setback. Furthermore, it facilitates isolation of that setback making it possible to monitor the students' response to the difficult passage.

Latin thus seemed to provide a particularly relevant setting to study how students are affected by difficulties and to what extent they bounce back from a setback. There also seems a possibility for (Dutch) Latin teachers to specifically teach their high ability students how to bounce back and learn from academic setbacks. Previous Latin instruction research focussed on domain specific proficiency and differences between students based on their proficiency. These existing studies mainly study the question what makes a good translator or what mistakes students make (e.g. Futch, 1935; Eikeboom, 1970; Sarkissian, 2008;

Florian, 2015; Karten, 2015; Boyd, 2018; Luger, 2018). Besides opportunities regarding academic buoyancy, there thus is also an opportunity for the current study to investigate the role of other student variables than proficiency within the Latin translation domain.

## 4 INDIVIDUAL DIFFERENCES

Learning is affected by individual differences. According to Sullivan (2009) students' learning processes are affected by their personal traits, the sum of which lead to their personal identity. Examples of individual differences include learning ability and attribution strategies (Williamson, 2018, p. 1-10). This led us to hypothesise that individual differences might moderate the effect of a setback on the student's translation processes. We, therefore, took three individual differences into account in our study: [1] Latin translation proficiency, [2] mindset preference and [3] frustration tolerance.

After having measured these individual differences, we included them as learner variables in our analyses to investigate to what extent they influence academic buoyancy. In the coming sections we introduce each of the three learner variables we included to represent individual differences between students.

### *4.1 Latin translation proficiency*

Domain specific proficiency is a student variable that we felt was likely to affect how students react to difficult tasks. In our case, domain specific proficiency was specified as Latin translation proficiency.

According to the Flow Theory of Csikszentmihalyi (1975), when a task is challenging but attainable students are most engaged. Task engagement is best when the difficulty of the task is neither too challenging nor too easy. In their study in which the expectancy of the task difficulty was manipulated, Hom & Maxwell (1983) concluded that task motivation and the amount of effort put into a task are influenced by the extent of the success or failure the student *expects* based on the perceived task difficulty. Proficiency of the task at hand thus seems to play a role in the extent students engage with a task when it is challenging.

Our target group of participants consisted of gifted students. Being academically gifted or a high potential does not automatically mean that the students are also all good Latin translators. Thus, proficiency was relevant to this study as a potential moderator of the effect of a setback.

### *4.2 Mindset*

Mindset is a term that Dweck has given to the system of beliefs held on to by students regarding intelligence. In general, it is thought that this system affects how students view and approach challenges (Yeager & Dweck, 2012). In the current study we include mindset to answer whether this also holds true for setbacks within a Latin translation context with gifted students.

Dweck distinguished a growth and fixed mindset (e.g. Dweck, 2011). On the one extreme of the spectrum there are students with a growth mindset. These students believe that intelligence is pliable and that challenges provide learning opportunities. On the other extreme, there are students with a fixed mindset, who believe intelligence is something static. Mindset seems domain specific and the mindset scale is a continuum, with students not only at either ends of the spectrum, but also anywhere in between.

The task processes of students with a more growth mindset and those with a more fixed mindset differ in various aspects. When students with different mindset preferences are compared to each other, studies have shown that students who are inclined to a growth mindset reflect more upon their work (Nussbaum & Dweck, 2008), which is an important feature of metacognition. According to multiple studies such as Cury, Da Fonseca, Elliot & Moller (2006) and Mueller and Dweck (1998), they also display more intrinsic motivation than students with a fixed mindset preference and studies have shown that adopting a growth view provides long term effects upon persistence and performance. Moreover, growth mindset holders are more equipped to dealing with disappointment due to having honed more effective coping strategies (Hong, Chiu, Dweck, Lin & Wan, 1999). In other words, when feeling unable to do something, fixed students are more likely to give up on the task.

Dweck (2012) seems to imply that when a student is gifted, the student is vulnerable towards fixed thinking and the problems related to it such as fear of failure or avoidance behaviour. Ziegler & Stoeger (2010) found support for their proposal that problems caused by a fixed mindset might only occur when students relate failure to their own shortcomings. However, more research is needed to determine how mindsets work in gifted students, despite the existing body of literature on this topic (McCoach & Flake, 2018, p. 208).

In sum, we included mindset as a learner variable for two reasons. First, the students' beliefs about difficult tasks might be relevant to the extent in which they bounce back from the difficult task, depending on whether they see challenges as positive or negative. Second, as our participants were gifted, this provided an opportunity to further explore the relationship between giftedness and mindset.

### *4.3 Frustration tolerance*

Frustration tolerance is a trait that affects the ability to overcome hurdles and endure demanding environments. It is connected to self-regulation. Irritability and refusing to participate are possible expressions of having a lower frustration tolerance (Bouman, 2011). If a student has low frustration tolerance, they are more at risk of avoiding challenges and thus not learning healthy strategies to deal with them (Solomon & Rothblum, 1984; Bridges & Roig, 1997; Harrington, 2005; Wright, Lam & Brown, 2009).

According to Silverman (2002, p. 31), gifted students are at particular risk of having a low frustration tolerance for their age as they are more used to instant gratification. The empirical observations of Grobman (2006) confirmed this. Gifted students being prone to low frustration tolerance, combined with the fact that low frustration tolerance might affect how students' approach (or even avoid) a difficult task made us include it as a final learner variable for our study.

## 5 THE PRESENT STUDY: AIMS AND RESEARCH QUESTIONS

Our main research aim was to establish how a lack of academic buoyancy in gifted students is reflected in their task process whilst translating Latin. We were also interested in determining to what extent the effect of a difficult task on their task processes is moderated by the learner variables translation proficiency, mindset and frustration tolerance.

This dissertation consists of three parts. The first part presents the instrumentation and methodology behind the studies, whereas the second focusses on answering the research questions. The final part forms the discussion of both the first and second parts. Below, we first present an overview of the structure of this dissertation and the questions we answer in the chapters. Then we summarize what can be found in each chapter.

Part I Instrumentation and methodology

| | |
|---|---|
| Chapter 1: | Which instruments did we use to gather data for our study and how did we develop them? |
| Chapter 2: | How did we select our participants and how did we implement the instruments to collect the data for the study? |
| Chapter 3: | How did we create the coding schemes necessary for analysing our think aloud data and what were the coding procedures we followed? |

Part II Academic Buoyancy and Latin Translation

Chapter 4: 1. To what extent are the participants' translation quality and processes affected by the setback?
2. To what extent is that the effect of the setback on the students' translation quality and processes moderated by their translation proficiency, mindset preference and frustration tolerance?

Chapter 5: To what extent do translation proficiency and mindset preference account for differences between the participants' display of metacognitive and affective strategies throughout the setback task?

Chapter 6: How do the learning strategies and patterns of academic buoyant students compare to those of non-academic buoyant students when they are faced with a task of a heterogenous complexity?

Chapter 7: To what extent was our research method valid and were our findings reliable?

*Chapter 1: Which instruments did we use to gather data for our study and how did we develop them?*

We developed a Latin translation task that triggered the students to experience a setback: the Latin Buoyancy Task. This task existed of three subsequent tasks. Two of them were relatively easy tasks which functioned as the pre- and post-measurements. The third part was visually similar yet unsolvable. This task served as the setback and the intervention. The creation of the Latin Buoyancy Task had the form of a design study which consisted of three design cycles. In the first part of Chapter 1 we present the underlying design principles and the subsequent design cycles of the Latin Buoyancy Task which led to the instrument itself.

In this first chapter we also present the development of four other instruments. The first of these was a task that we used to purposefully select participants for the study. This instrument was also created in the form of a design study. This design study is the central topic of the second part of the first chapter.

The third part of Chapter 1 relates to the development of two questionnaires to measure the learner variables. First, we created a Dutch version of a three-itemed questionnaire by Dweck, Chiu and Hong (1995) that has often been used in studies to measure mindset (e.g. Blackwell, Trzesniewski, & Dweck, 2007; Romero, Master, Paunesku, Dweck & Gross, 2014; Yeager et al., 2019). Previously, no Dutch version of this questionnaire had been statistically validated. Secondly, we also translated, adapted and validated a second questionnaire aimed at measuring frustration tolerance (Wright, Lam & Brown, 2009). The development of these questionnaires is presented in the second part of Chapter 1. The fourth final part of the first chapter focusses on a measurement we created to specifically collect data retaining to affective learning processes. All in all, the first chapter includes a complete overview of all the instruments we developed to collect the data for our study.

*Chapter 2: How did we select our participants and how did we implement the instruments to collect the data for the study?*
In Chapter 2 we focus on how we implemented these instruments to collect the data for our study. We first introduce why and how we made use of extreme sampling to purposefully select our participants. We then briefly introduce the 16 students that participated in the study as presented in the later chapters. These 16 participants performed the Latin Buoyancy Task while thinking aloud. We present the arguments for this method of data-collection with the addition of retrospective thinking aloud in Chapter 2. These complementary data sets provided us with insights into their thoughts and behaviour. Thus, we were able to monitor differences in their outcome in terms of translation accuracy and task process before, after and during the setback task.

*Chapter 3: How did we create the coding schemes necessary for analysing our think aloud data and what were the coding procedures we followed?*
Collecting data via think aloud methods necessitated us to make use of a coding scheme to analyse the verbal transcripts. As no coding scheme existed which suited our purposes, we developed a coding scheme, the aim of which was to be able to code the thinking aloud data thoroughly. We also created a derivative coding scheme, which allowed for more room in the coding for affection, based on context and the data gathered via the retrospective thinking aloud. The main coding scheme and its condensed derivative gave us flexibility in coding, depending on the research question at hand. Both the development of the coding schemes and the coding process itself is are presented in Chapter 3.

*Chapter 4: To what extent are the participants translation quality and processes affected by the setback and to what extent is that the effect of the setback on the students' translation quality and processes moderated by their translation proficiency, mindset preference and frustration tolerance?*

The aim of Chapter 4 was twofold. First, we aimed at ascertaining whether and in which way the participants' translation quality and processes were affected by the setback. To do so, we compare and contrast the accuracy of the translations of the pre- and post-setback tasks, alongside three process variables -implicit metacognition, perseverance and grit. Our second aim was to describe to what extent that effect was moderated by the learner variables. The analysis of the data for this explorative study was done quantitively.

*Chapter 5: To what extent do translation proficiency and mindset preference account for differences between the participants' display of metacognitive and affective strategies throughout the setback task?*
Chapter 5 specifically focuses on the setback task (i.e. the intervention) itself. The aim of this study was to describe how translation proficiency and mindset preference account for differences between the participants' display of metacognitive and affective strategies throughout the setback task. In this study, the analysis of the think aloud data was qualitative. Our focus in this chapter only lied on the setback task and the included learner variables, which accounted for individual differences between the participants.

*Chapter 6: How do the learning strategies and patterns of academic buoyant students compare to those of non-academic buoyant students when they are faced with a task of a heterogenous complexity?*
Chapter 6 forms the penultimate chapter of this dissertation and the presents the final study. In this chapter we qualitatively delve deeper into the verbal transcripts, in contrast to Chapter 5, we do not only focus on the setback task, but on the process of the students throughout the whole Latin Buoyancy Task. We aim to create student profiles that differentiate between learning strategies and patterns of learning strategies that are specific to the students that do bounce back after the setback and those who do not.

*Chapter 7: To what extent was our research method valid and were our findings reliable?*
In the seventh and final chapter, we reflect on the study and present a general discussion of the study as a whole.

# PART I

# INSTRUMENTATION & METHODOLOGY

# CHAPTER 1

# MEASURING INSTRUMENTS

The academic buoyancy of students with high cognitive potential is the central topic of this dissertation. Our main aim was to establish how a lack of academic buoyancy in gifted students is reflected in their task process. To do so, we posed the question: How do individual differences regarding translation proficiency, mindset preference and frustration tolerance affect the extent that the students bounce back after a difficult task? In the current chapter we present the four instruments that were developed to realize our research aims regarding academic buoyancy. Each instrument went through its own design process and was tested separately. The design process of two of the instruments consisted of the adaptation of existing English questionnaires. The other two instruments were designed from the ground up, specifically for the current research project.

We used a design research method to develop a task-based instrument to trigger an academic setback: the Latin Buoyancy Task. The design process of this task is presented in Part I of this chapter. Part II focusses on the task used to select the participants. Similarly to the Latin Buoyancy Task, it aimed at creating an academic setback, albeit in another linguistic domain. As with the other instruments introduced in this chapter, not only the selection task itself is presented, but also the different phases in which it was tested before being used for the central study of this dissertation. In Part III we present the two questionnaires which were aimed at collecting data on the student characteristics included in this study. Finally, the design of the instrument intended to measure the participants' affective state during the task is the subject of Part IV.

# 1 INTRODUCTION

Martin & Marsh (2008) defined academic buoyancy as 'the ability of students to successfully deal with academic setbacks and challenges that are typical of the ordinary course of school life' (p. 54). Examples of such setbacks include being faced with difficult tasks, receiving isolated poor grades and demanding deadlines. Over the last decade the body of literature relating to academic buoyancy has consistently grown (e.g., Malmberg, Hall & Martin, 2013; Kim & Han, 2014; Comerford, Batteson & Tormey, 2015; Smith 2015; Jahedizadeh, Ghonsooly & Ghanizadeh, 2019).

In this growing body of studies, a set of self-report questions is the main method of measuring the extent of buoyancy in students. However, generally, self-reports by adolescent students does not always provide the most valid data as adolescents are inclined to giving socially desirable answers (Fan, et al., 2006). For academic buoyancy specifically, doubts have also been expressed about the validity of the data collected by self-report. To avoid problems with self-report, Verrier, Johnson & Reidy (2018) proposed the Teacher Academic Buoyancy scale, a measurement filled in by teachers about the buoyancy in their students. However, they concluded that their teacher driven measure was not viable as teachers are unaware of student behaviour that is indicative of academic buoyancy. This is in line with Zsu & Urhahne (2014) who found that teachers are unable to correctly judge students' negative feelings. Thus, it seems other methods than self-report or teacher evaluations are needed to provide more insights into academic buoyancy.

An option for such a method seems to be prompting failure under lab conditions. Lucas, et al. (2015) isolated failure for example by using unsolvable items and studied behavioural differences between gritty and non-gritty students, leading them to determine how persistence is affected by grit. Isolating failure thus seems a viable method to gain insights into task behaviour. As failure at a task is an example of an academic setback, we hypothesized that isolating failure would also be viable for studying academic buoyancy. We assumed that differences between buoyant and non-buoyant students would be visible in the outcome of a task which was performed directly after being confronted with failure: the quality of buoyant students' work would not decrease, where that of non-buoyant students would significantly decline. We based this hypothesis on the premise that the confrontation with failure would lead the non-buoyant students to negatively adapt their task-solving strategies. To test these assumptions, a task-based measure needed to be developed.

We therefore aimed at creating a task that prompted failure and measuring the extent of academic buoyancy in students through the task. We operationalised academic buoyancy by task accuracy. Therefore, this instrument would have to consist of three subtasks: a pre-setback task (i.e. the baseline task), a challenging subtask functioning as the academic setback and a post-setback task. Such a design would allow us to isolate 'failure' and compare the participants' accuracy pre- and post-setback. In turn, by having the students perform the task whilst thinking aloud, this difference could be compared to what happened to their task processes throughout the three subtasks. Thus, we set out to create a three-part task.

We settled that translating Latin would provide a good domain for a task aimed at researching academic buoyancy, for reasons explained in the Introduction (p. 6-8). We decided that each subtask would consist of three Latin sentences (long enough for the outcome not to be a fluke, but short enough to be able to do three sets within a reasonable amount of time). If such a task indeed was helpful in collecting data on academic buoyancy, it could be used to establish how a challenging linguistic task (i.e. a temporary academic setback) affects gifted students' use of cognitive, metacognitive and affective strategies and to answer how the student characteristics domain specific proficiency, mindset preference and frustration tolerance affect the extent the students would bounce back after a difficult task.

To ensure the translation task matched our aims, we established multiple design principles. The first was that the second task would lead to the participants certainly experiencing a setback. Further, to ensure that the pre- and post-setback tasks could be compared, they had to be of equal difficulty and individually internally consistent. This led to the first three design principles:

1. The second sub-task must ensure that the students experience a setback, similar to setbacks they could experience at school under normal circumstances;
2. The pre- and post-setback tasks must be internally consistent;
3. The pre- and post-setback tasks must be of similar difficulty.

To guarantee that the participants would experience a setback, the second subtask had to be noticeably challenging for the participants. According to Sasayama (2016) students only notice differences in the complexity of linguistic tasks when the difference is very large. To ensure that all the participants, regardless of their Latin abilities would notice the difficulty increase and experience it as a setback, we decided that the Latin should be syntactically nonsensical and, therefore, untranslatable. However, the participants needed to experience it as a setback. They should not have the impression of being presented with a strange task or even guess it was an intervention with a specific goal, in contrast to a 'normal' translation task. The task must neither be obviously nonsense Latin nor too different from the other subtasks. This is why the sentences contained real Latin words and morphological forms, but a non-sensical syntax.

Moreover, the participants needed to feel they were able to complete the first task, so they could provide us with information on how they approach a translation task that is not too much of a challenge to them. However, if the task was too easy, the participants might not need to apply many problem-solving strategies. Furthermore, a task-based instrument provided the possibility for exploring academic buoyancy quantitively and qualitatively by letting the participants perform the task thinking-aloud. For successful thinking-aloud the task needed to be complex enough to facilitate different strategies (Ericsson & Simon, 1980; Schellings, 2011; Veenman, 2011). We, therefore, added two design principles:

4. All the subtasks must feel familiar to third year Latin students. This particularly holds for the second subtask, despite the possible translations being in no way satisfying or leaving the students feeling they might be correct.
5. The pre-setback task must feel attainable to students, yet still be complex enough to encourage students to use (cognitive, metacognitive and affective) problem-solving skills and make thinking-aloud possible.

The doable-undoable-doable pattern of the subtasks in itself might feel familiar to students as it is actually an extreme simulation of the heterogeneous nature of the Latin texts that students are expected to translate, where complex sentences can be followed by much easier ones. Besides being manipulated to include untranslatable items, a difference between those Latin texts and our task is that our task consists of unconnected sentences that do not form a story. However, Latin students are used to such exercises as they are frequently asked to translate independent sentences when learning new grammatical elements. Thus, we expected the Latin Buoyancy Task to feel familiar to students.

A final design principle was derived from prior studies related to the translation process of Latin students. Multiple of these studies noted the negative effect that vocabulary deficiencies have on the students' translation process (e.g. Van Krieken, 1982; Sarkissian, 2008; Florian, 2015). For our study, vocabulary deficiencies might have led to participants being inadvertently challenged in the pre- and post-setback tasks, thus diminishing the effect of isolating the setback. Therefore, the sixth design principle was formulated as:

6. The task must ensure that the participants are not hindered by a lack of vocabulary knowledge.

To ensure our instrument met the final design principle, we decided that each subtask would be accompanied by a complete vocabulary list of the words that occurred in the sentences.

In sum, the aim was to study academic buoyancy by developing a task-based instrument, which included an academic setback. We hypothesized that not all participants would be able to bounce back to their translation accuracy and strategies from before the setback and therefore the instrument could be helpful in acquiring data on the process of academic buoyancy. The basis for this instrument would be a Latin translation task consisting of three sub-tasks following a doable, undoable doable pattern and adhering to the six design principles. As it happened, during the design of the instrument, we not only found that in the post-setback task some students indeed did not bounce back to their pre-setback translation accuracy, but also that a group of students significantly improved upon their pre-setback accuracy.

## 2 THREE DESIGN CYCLES

In the following sections we present the design research that led to the final version of the Latin Buoyancy Task. Each of the three design cycles consisted of a round (re)designing the task followed by a testing and feedback round with different third year Latin students. The outcomes were then analysed before starting a new cycle or finalizing the instrument. The results found from the analysis of the third cycle led us to conclude that the instrument was ready to be used for a next phase of the study, thus no fourth cycle was started. We then piloted the instrument with a student performing it whilst thinking aloud (see Chapter 2, p. 89). In the coming sections we present the method, results and discussion of each design cycle. The cycles are summarized in Table 1.8 at the end of the discussion of the third and final design cycle.

## 3 DESIGN CYCLE 1

### *3.1 Method*

The aim of the first design cycle was to compile two groups of three Latin sentences that would function as the pre- and post-setback tasks. By the end of this cycle, we aimed to be able to fulfil design principles 5 and 6, in full and design principle 4 for the pre- and post-setback tasks.

### 3.1.1 Participants

In the first design cycle, 28 students participated. The participants of the first design cycle formed one third-year Latin class. Their ages ranged from 14 to 15 years old. They attended the researchers' own school and were students she taught. This school was a so-called *scholengemeenschap,* where Latin is not compulsory. Prior to participating in the design study, the students' parents were informed of the study and asked for passive consent. No refusals of participation in the study were received.

### 3.1.2 Materials

To construct the Latin Buoyancy Task, we needed six Latin sentences that were translatable for third-year students, to function as the pre- and post-setback tasks. These six translatable Latin sentences would eventually have to be split into two more or less similarly complex groups of three sentences. To start the first design cycle, we selected twelve Latin sentences from works by classical Roman authors, that we deemed translatable for most third-year students based on the grammatical elements found in the sentences. These sentences were found in the exercises and examples provided by Koenen (2007) and Kroon (2007). The twelve sentences were semi-randomly turned into four sets of three, whereby, where possible, each set contained a sentence with a ditransitive verb with at least one other element such as an attribute or a satellite. Each set also contained two sentences including a tritransitive verb and an attribute or satellite. This led to the following four sets of sentences found in Table 1.1.

### 3.1.3 Procedures

Design cycle 1 took place over four Latin lessons. At the start of the first lesson, the researcher/teacher explained that she was researching how students translate Latin sentences and that the students would be helping her design of the necessary instruments by completing some tasks in the coming lessons. The students were told they were to translate three sentences from Latin into Dutch, using a wordlist. It was emphasized that it was not a test of their abilities and all analysis would be done anonymously. They were also asked to keep note of the time in the designated spots and fill in any comments they had about the task sentences and wordlist on the answer sheet. To help them do so a timer was projected on the board.

After checking whether the students had any questions, the task was distributed upside down and the students told to leave it so until otherwise informed. When all students had the task, they were told to turn the paper over and translate the sentences. After 10 minutes all students had finished translating the three sentences and had written down their comments. A short post class discussion then took place led by the researcher/teacher. The researcher/teacher asked the class which sentences they found easy, difficult and why they thought that. Overall, the general feeling expressed was that the sentences were difficult to translate, and students found the wordlist helpful. After the class discussion the students handed in their answer sheets. This process was then repeated in three further lessons and no significant incidents occurred during the data-collection.

*Table 1.1 Sets of Latin sentences compiled and tested in the first cycle.*

| | Sentence 1 | Sentence 2 | Sentence 3 |
|---|---|---|---|
| Set 1 | Tridente suo terram percussit.[1] | Roscius praedia aliis coluit.[2] | Octavianus armis expulit ex urbe collegam.[3] |
| Set 2 | Veniam ille amori forsitan nostro dabit.[4] | Caesar suos a proelio continebat.[5] | Equus pilo traiectus est.[6] |
| Set 3 | Gladiator in arena consilium capit.[7] | Otium, Catulle, tibi molestum est.[8] | Equitatum auxilio Caesari Haedui miserant. [9] |
| Set 4 | Vitam iuncundissimam vivo.[10] | Sol cuncta sua luce complet.[11] | Sacra tibi commendat Troia penates.[12] |

[1] Ov. *Met.* 1.283. He made the ground tremor with his trident.
[2] Cic. *S. Rosc.* 49. Roscius cultivated the farms for others.
[3] Cic. *Cat.* 3.24. Octavianus expelled his colleague from the city with weapons.
[4] Sen. *Ph.* 225. Maybe he will give our love forgiveness.
[5] Caes. *DBG.* 15. Caesar kept his men away from the battle.
[6] Liv. *AUC.* 27. The horse is impaled by the spear.
[7] Sen. *Ep.* 22.1. A gladiator comes to a plan in the arena.
[8] Cat. *Carm.* 51. Idleness is troublesome for you, Catullus.
[9] Caes. *DBG.* 18. The Haedui had sent the cavalry with help to Caesar.
[10] Sen. *Ep.* 32. I live a very pleasant live.
[11] Cic. *Rep.* 6.17. The sun fills everything with his light.
[12] Verg. *Aen.* 2.293. Holy Troy entrusts its Penates to you.

### 3.1.4 Analyses

After the four lessons had been completed, the students' translations were scored for accuracy by two assessors, who were both acquainted with rating Latin translation tasks for accuracy. To rate the accuracy, we made use of a three-point scale. A translation that was morphologically and syntactically correct received 2 points, a translation containing one error received 1 point and translations with more than one morphological error received 0 points. An example of the scoring system can be found in Table 1.2. After comparing the students' scores as given by the two assessors no inconsistencies were found.

To compare the difficulty of the sentences, we compared the means of the scores from each sentence. We had asked the students to make note of the time at designated parts of the task. However, most of the students had left the timestamps empty. Thus, due to their incompleteness we did not analyse these.

*Table 1.2 Scoring system for the Latin Buoyancy Task*

| *Octavianus armis expulit ex urbe collegam.* | | |
|---|---|---|
| Score | Explanation | Example |
| 2 | Correct translation | Octavianus expelled his colleague from the city with weapons.[1] |
| 1 | Nearly correct: One error | Octavianus expelled his colleague from the city with <u>a weapon</u>.[2] |
| 0 | Multiple errors | Octavianus <u>expels</u> his <u>colleagues</u> from the city with weapons.[3] |

## *3.2 Results*

In Table 1.3 we present the mean scores for the sentences' accuracy. As Table 1.3 demonstrates, students only scored an average of at leaste 1.0 points on three sentences (*Octavianus armis expulit ex urbe collegam; Equus pilo traiectus est; Vitam iuncundissimam vivo).* The students had translated the other sentences quite poorly.

This was confirmed in the class discussion: the students said they found nearly all the sentences very difficult to translate, even with the supplied wordlists.

[1] *Octavianus verdreef met zijn wapens zijn collega uit de stad.*
[2] *Octavianus verdreef met een wapen zijn collega uit de stad.*
[3] *Octavianus verdrijft zijn collega's met wapens uit de stad.*

Students also said they felt that some of the sentences seemed to have a vague or mainly philosophical intention. The students also said that they kept forgetting to write down how long they had spent translating each sentence as they were too focused on the task itself. Going through the written feedback by the students, there were some comments relating to students not understanding the notes 'I thought Roscius meant 'a wealthy Roman landowner'[1] but it was just a name and the missing odd word from the wordlist. However, most of the feedback implied that the sentences were too difficult or not understood by students (e.g. 'weird sentences[2]).

*Table 1.3 Mean scores of the translations of the tested sentences*

| | Set 1 | | | Set 2 | | | Set 3 | | | Set 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sentence | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| *M* | .52 | .29 | 1.41 | .92 | 1.67 | .72 | .21 | .36 | .42 | .78 | 1.59 | .33 |
| *SD* | .28 | .91 | .43 | .81 | .19 | .12 | .11 | .32 | .24 | .39 | .13 | .21 |

### *3.3 Discussion*

For the second design cycle, we needed six sentences which had proven to be translatable for the students, but still provided challenges (design principle 5). These six sentences would then need to be split into two groups of three sentences that were of comparable difficulty to act as the pre- and post-setback task (design principle 3). However, only three sentences were scored decently on (*Octavianus armis expulit ex urbe collegam; Equus pilo traiectus est; Vitam iuncundissimam vivo).* The difficulty of the sentences was confirmed in the classroom discussion and the students' written notes. We were thus unable to select six translatable sentences that could be divided into two groups of comparable difficulty and design principle 5 had not been met. The sentences did look and feel like a task they were used to performing in Latin lessons, therefore, design principle 4 was met for these sentences to some extent. Finally, seeing that there were some words missing from the vocabulary list also the sixth design principal was not fully met.

---

[1] *Ik dacht dat Roscius gewoon een rijke Romeinse grondbezitter betekende.*

[2] *Rare zinnen*

Asking the third-year participants to translate only authentic Latin, thus, seemed a bridge too far. To solve this problem, we decided to include the three sentences that had been translated relatively well in the next design cycle. We would then supplement these with adapted versions of three other sentences. To decide which three sentences would be adapted and how they were to be adapted, we would start the second design cycle by going through all the translations again and searching for common errors and looking for possibilities to prevent these common errors with small changes to either the wordlist or the Latin itself.

# 4 DESIGN CYCLE 2

## *4.1 Method*

The purpose of the second design cycle was twofold. First, we aimed at improving our findings from the first design cycle regarding the pre- and post-setback tasks. After the first design cycle, we did not yet have enough sentences that felt attainable yet not too easy to third-year students (design principle 5). Also, design principle 6 had not yet been met. Thus, our first aim of the second design principle was to meet these design principles. Further, we intended to create three non-sensical Latin 'sentences'. These would be tested for them indeed being syntactical nonsense, whilst not appearing to be untranslatable to the students and, thus, creating a sense of failure (design principles 1 and 4).

### 4.1.1 Participants

In the second design cycle, a total of 28 participants participated. 26 of the participants completed all parts of the design cycle. The participants were in their third-year of Latin and their ages ranged from 13 to 15 years old. These participants attended the same school as the participants from the first design cycle, however, they were from a different class and were not taught by the researcher. Prior to participating in the design study, the students' parents were informed of the study and asked for passive consent. No refusals of participation in the study were received.

### 4.1.2 Materials

We made use of six Latin sentences. As presented at the end of the first design cycle, we included the three sentences that had been proven to be translatable for third-year students. For the other three sentences, we went through the translations errors to find frequently occurring problems, which might easily be averted with a small change to either the sentence or the wordlist. Examples of such problems included interpreting *tridente* as the subject of *percussit;* not translating *dabit* as the future tense and not translating *suos* as a plural object of *continebat*. Based on these findings we made the following three changes to the sentences:

1. We added *Neptunus* as an explicit subject resulting in: *Neptunus tridente suo terram percussit;*
2. We changed the future tense of *dabit* to the present resulting in: *Veniam ille amori forsitan nostro dat;*
3. For *suos* in *Caesar suos a proelio continebat* we added a note to the word list saying that *suus* in this sentence meant 'his man'.

These new sentences were added to the three that we had found to be in the first design cycle. Figure 1.1 presents an overview of the six sentences tested in the second cycle.

*Figure 1.1 Sentences tested in the second design cycle.*

1. Octavianus armis expulit ex urbe collegam;
2. Equus pilo traiectus est;
3. Vitam iuncundissimam vivo;
4. Neptunus tridente suo terram percussit;
5. Veniam ille amori forsitan nostro dat;
6. Caesar suos a proelio continebat.

For the second part of this design cycle, we also needed 'sentences' that would function as the setback task. To create the untranslatable Latin, we combined Latin words, sometimes within the same word field to make the sentence seem possible at first glance. Despite some words that might be used in a sentence together, syntactically and lexically the words do not form a meaningful whole. The words were then given morpho-syntactical endings, however, in each sentence at least one element was included that could not be incorporated in any manner with the rest of the elements from the sentence. The morpho-syntactical endings were all real, to not be too obviously impossible. The sentences which were meant to form the setback, including examples of what makes them untranslatable are found in Figure 1.2.

*Figure 1.2 The setback, i.e. untranslatable task.*

| 1. *Hic* | *sine* | *filium* | *longe* | *gravissimum* | *cano.*[1] |
|---|---|---|---|---|---|
| Here | without<br>+ablative | the son<br>accusative | long<br>adverb | Very heavy<br>accusative | I sing |

| 2. *In mare* | *habes* | *Delphinus* | *et* | *canorum* | *promittere.*[2] |
|---|---|---|---|---|---|
| In the sea<br>ablative | you have | Dolphin<br>nominative | and | of dogs | to promiss<br>+ accusative /<br>infinitive |

| 3. *Cur* | *esse* | *ea bona* | *periit* | *similis* | *copia.*[3] |
|---|---|---|---|---|---|
| Why | to be | those things<br>nominative /<br>accusative | he perished | resembling<br>+ genitive | reinforcements |

Examples of incoherent syntax per sentence:
1. *Sine* (without) necessitates an ablative, no ablative can be found in the rest of the sentence
2. *promittere* (promiss) necessitates an accusative or second infinitive, neither can be found in the rest of the sentence.
3 *esse* (to be) is an infinitive, however, in the rest of the sentence no grammatical reason for this infinitive can be found.

### 4.1.3 Procedures

The conditions for the data-collection were similar to that of the first cycle. As the researcher was acquainted with the students, despite not being their teacher, no introductions were made. Their regular teacher was present in an audience role. The researcher explained that she was researching how students translate Latin sentences and that the students would be helping her design the instruments by completing the following translation task. The students were told they were to translate six sentences from Latin into Dutch, using a wordlist. They were also asked to fill in any comments they had about the task sentences and wordlist on the answer sheet. It was emphasized that it was not a test of their abilities and all analysis would be done anonymously. A change compared to the first cycle, was that we did not ask the students to make a note of how long they had spent translating each sentence, as we had found that the students generally forgot to do so.

After checking whether the students had any questions, the task was distributed upside down and the students told to leave it so until otherwise informed. When all the students had the task, they were told to turn the paper over and translate the sentences. After 20 minutes all the students had finished translating the six sentences and had written down their comments. This was followed by a short class discussion, led by the researcher. The researcher asked the class which sentences they found easy or difficult and why they thought so.

Overall, the students were more positive about their ability to translate the sentences than the students from the first design cycle were. After the class discussion the students handed in their answer sheets. No significant incidents occurred during the data-collection.

In a separate session this group of students, including two who were previously absent (making $N$ = 28), was also the first test group of the untranslatability of the three incoherent Latin sentences. The aim of testing the incoherent sentences was to establish that students could not create logical sounding translations, as this would diminish the effect of the setback (see design principle 1). This was done under similar conditions as the testing of the translatable sentences: the students were placed in test setup, their teacher and the researcher were present, they were supplied with something that to all sense of purposes were three Latin sentences, an accompanying wordlist and a place to write feedback on the sentences. Again, the researcher explained that she was researching how students translate Latin sentences and that the current task would provide her with information needed to develop translation tasks. It was emphasized that it was not a test of their abilities and that the analysis would be done anonymously.

The task was distributed face-down and the students were told to leave it so until otherwise informed. When all the students had the task, they were told to turn the paper over and translate the three sentences. After 20 minutes, (i.e. as long as they needed for six translatable sentences), all the students were told to finish translating the three sentences and told to write down their comments. Prior to this most students had stopped translating. Based on Lucas et al. (2015), it seemed important to stop the students at some point from continuing as the grittiest students might not come to that conclusion themselves and the current aim was to test the untranslatability of the sentences, not the grittiness of the students.

The translation task was followed by a short post class discussion, led by the researcher. The researcher asked the class which sentences they found easy or difficult and why they thought that. The students were very vocal about it being too difficult. The researcher then told the students that they were right and that the task was not even translatable. Most students started laughing or said to be relieved it was not down to their lack of abilities. A single student seemed frustrated by being given an untranslatable task. The researcher then explained that Latin texts are of an heterogenous nature and that the research was meant to investigate whether and how students are able to bounce back after a difficult translation passage. This peaked their interest and some students started telling the researcher if they thought they were students who were negatively affected by challenges or not. Besides certain behaviour (movement, sighing, pen ticking) that was seen in some students during the untranslatable 'sentences', no significant incidents occurred during the data-collection.

### 4.1.4 Analyses

The same three-point rating system was applied to the translations of the six translatable sentences as in the first cycle (see Table 1.2). As in the first cycle, we compared the means of the scores from each sentence to ascertain the difficulty of the sentences.

For the untranslatable 'sentences', the students' translations were not scored as there was no actual solution, but each translation was checked to see if the Dutch result made any sense. If there were students with logical sounding translations the sentences would not be deemed incoherent enough to adhere to design principle 1. If a student thought their translation was correct, due to it sounding logical, the student would not have experienced a setback, and in that case the subtask would not fulfil the design principle of 'the second sub-task must ensure that the students experience a setback'.

## *4.2 Results*

The scoring now increased significantly compared to the first design cycle. On average the sentences were now all scoring above one point out of two. The average scores per sentence can be found in Table 1.3. Compared to the first cycle, the classroom discussion also was more positive regarding the translatable sentences.

*Table 1.3. Scores per sentence in the second cycle (N = 26)*

| | Sentences | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| *M* | 1.15 | 1.38 | 1.62 | 1.35 | 1.42 | 1.19 |
| *SD* | .83 | .64 | .70 | .56 | .81 | .80 |

*Note. The maximum score per sentence was 2.*

Regarding the untranslatable sentences, none of the students' attempts proved to make sense. Moreover, the relief seemed to imply that the students were not aware that it was untranslatable. This was confirmed by the students' written feedback, which often included phrases as 'I know this is wrong, but I cannot do any more.'[1].

### *4.3 Discussion*

The new average scores implied that, generally, the six sentences now were translatable for third-year Latin students, but that there were still some difficulties in the sentences. These sentences, therefore, seemed to fit design principle 5. We thus decided that we would continue to the next design cycle, i.e. seeing whether we could create two equivalent sets, with the current versions of the sentences. We also improved wordlists.

The lack of sense or meaning found in any of the students' attempts to translate the non-sensical, combined with the verbal and written remarks on the difficulty of the 'sentences' seemed to indicate that the first design principle had also been met. None of the students had realised that the 'sentences' were actual nonsense. We interpreted the relief the students' expressed when finding out that there was no actual translation possible, as an indication that they had thought that they should be able to complete the task (design principle 4).

In both cases of the translatable and non-translatable sentences there were no written comments related to the wordlists. Nor did the students mention them during the classroom discussion. We, therefore, determined that the sixth design principle had been met.

## 5 DESIGN CYCLE 3

### *5.1 Method*

Having tested the improved translatable sentences and having confirmed that the untranslatable sentences were indeed untranslatable, we moved on to the third design cycle. During this cycle we aimed to test the translatable and untranslatable sentences as they were intended to be used: three sentences working as a pre-setback measurement, the untranslatable sentences as the setback and finally the three remaining sentences functioning as a post-setback measurement. For the pre- and post-setback tasks, the aim was to combine the sentences into two sets of three, which were of comparable difficulty and internally consistent (see design principles 2 and 3).

[1] *'Ik weet dat het niet klopt, maar ik kan er niets beters van maken.'*

### 5.1.1 Participants

The final cycle originally included 87 participants, however for two of them the data was incomplete. Their results were not included in the analysis, leaving us with 85 participants. Again, the students participating in the testing of the Latin Buoyancy Task were in their third year of Latin. Their ages range from 14 to 16 years old. The participants attended three different schools. These were different schools than the schools of the students who participated in the first two cycles. Passive consent was received from the parents prior to participation.

### 5.1.2 Materials

We made use of the six Latin translatable sentences and three untranslatable 'sentences' as presented in the second cycle. To create two equal sets of three sentences to function as the pre- and post-setback tasks (see design principle 3), we sorted the six sentences into two near equal groups according to the average means as found during in the previous design cycle (see Table 1.3). These sentences were then ordered by complexity of the syntax from least complex to most complex. We based the order on complexity of the syntax and not on the mean as found in the second design cycle, so students would have to use Latin knowledge they previously had not yet needed in the task. We did this in the hope that the thinking-aloud would benefit, due to the task not only being solvable by making use of knowledge and strategies already used before. This led to the two tasks as presented in Table 1.4.

*Table 1.4. Two near equal sets of Latin sentences for the third design cycle*

| Set | Sentences | | Mean | *SD* |
|---|---|---|---|---|
| | Equus pilo traiectus est. | | 1.38 | .64 |
| A | Caesar suos a proelio continebat. | | 1.19 | .80 |
| | Veniam ille amori fortisan nostro dat. | | 1.42 | .81 |
| | | Total: | 3.99 | |
| | Vitam iucundissimam vivo. | | 1.62 | .70 |
| B | Neptunus tridente suo terram percussit. | | 1.35 | .56 |
| | Octavianus armis expulit ex urbe collegam. | | 1.15 | .83 |
| | | Total: | 4.12 | |

For the testing phase, the sets were made into booklets consisting of a pre-setback task, the setback task and a post-setback task. To ensure the researcher knew at a glance that the tasks were indeed being made in the right order by the students, each task had been printed on different coloured paper, (pre-setback: salmon; setback: lilac; post-setback: pink).

### 5.1.3 Procedures

The data was collected at three schools by the same researcher. Each time the data collection took place in a classroom, during a Latin lesson. The conditions were similar to the previous testing phases: The students' regular teacher was present in an audience role, the students were sat in test-setup and no significant incidents occurred during the data-collection. The researcher introduced herself and explained that she was doing a study on how students translate Latin sentences. The students would be testing translation tasks for the study. It was then emphasized that it was therefore not a test of their abilities and all analysis would be done anonymously.

Then, the randomly ordered booklets were distributed amongst the students, who were instructed to keep the booklets shut. To rule out an order effect, the booklets had been compiled making use of a switching replications method. Some students, therefore, received Set A as the pre-setback task and Set B as post-task, whereas for others this was inverted. When all the students had received a booklet, they were given instructions along the following lines: 'you have seven minutes to translate the three sentences on the salmon page, if you are finished before the seven minutes are up, you close your booklet. During and after translating the sentences you may not confer'. The students then set about the pre-setback translation task. After seven minutes, the students were told to close their booklets if they had not already done so. This process was then repeated for the setback task and the post-setback task.

After the three tasks had been completed, a classroom discussion followed. To start with, the researcher asked students to raise their hands if they thought the first set was the easiest. She asked the same question for the second and third set. In all the classes, most of the students thought the first set was the easiest. Moreover, all students indicated that the second task was the most difficult. After a short exchange of what they thought of the tasks, in which students mainly wanted to talk about the second set, the researcher informed the students that the second set was untranslatable. She also gave them an explanation about how students can react differently to challenges. During this debriefing, like during the previous testing phase, students were elated with this news. The most given answer to how does that knowledge make you feel was 'relief'.

### 5.1.4 Analyses

The students' translations of the pre- and post-setback tasks were rated for accuracy by the two assessors who had also previously rated the sentences in the first two design cycles. The rating system remained the same (see Table 1.2) and there were no inconsistencies in the ratings.

Statistical analysis in the third cycle was performed. To test for internal consistency (see design principle 2), only the pre-setback task was taken into consideration, as the untranslatable task was expected to affect the students' scores thereafter. An independent samples test was run to determine that Set A and Set B were similar in difficulty (see design principle 3).

To determine whether there was an effect of the setback, we first converted the scores into z-scores to make comparing between the two versions possible. Then we subtracted the scores of the post-setback task from the pre-setback task. Students whose scores had decreased more than a standard deviation compared to the pre-setback task were taken to be significantly affected by the setback. A t-test was run to determine if there was an effect of the untranslatable task and its significance, and Cohen's $d$ was calculated for the effect size.

## *5.2 Results*

The internal consistency of version A was .65 (Cronbach's alpha). As shown in Table 1.5, the students scored an average of three ($SD$ = 1.86) out of the available six points and the internal consistency would have been better if the first sentence was deleted.

*Table 1.5 The Latin Buoyancy Task scores of version A (N = 41)*

| Sentence | M | SD | Cronbach's alpha if item deleted |
|---|---|---|---|
| 1 | 1.02 | .76 | .831 |
| 2 | 1.12 | .81 | .216 |
| 3 | .85 | .85 | .436 |
| Total | 3.00 | 1.86 | |

The internal consistency of version B of the Latin Buoyancy Task was .72 Cronbach's alpha. As shown in Table 1.6, the students scored an average of 3.91 (*SD* = 1,43). Compared to version A, these sentences were translated slightly better. The difference, (-.91) was significant (*t* = (83) -2.54, *p* = .01): sentences from task B were easier to translate.

*Table 1.6 The Latin Buoyancy Task scores of version B (N = 44)*

| Sentence | M | SD | Cronbach's alpha if item deleted |
|---|---|---|---|
| 1 | 1.14 | .80 | .511 |
| 2 | 1.32 | .77 | .581 |
| 3 | 1.52 | .76 | .534 |
| Total | 3.91 | 1.43 | |

Regardless of version, the post-setback sentences tended to be less accurately translated than the pre-setback ones. We found that, on average, the version A students scored .15 points less (*SD* = 2.37) and the B version students 1.86 points less (*SD* = 2.17). Overall, around 35% of the students scored significantly (>1 *SD*) different after the setback than prior to it. Notably, of the 35% of the students that saw a significant change after the setback task, some of the scores of these students did not decrease, but actually improved. The distribution of the difference scores between the pre-setback and post-setback tasks is presented in Table 1.7.

A paired-samples t-test indicated that the difference between the first and third part of the task was significant, (*t*(84) = 3.95, *p* = <.001). The effect size was small according to the calculations using Cohen's *d* (*d* = .43).

*Table. 1.7 The distribution of the difference scores (N = 85)*

| | Significant Loss (>1 *SD* higher) | No Significant Change | Significant Gain (>1 *SD* lower) |
|---|---|---|---|
| Frequency | 16 | 53 | 16 |
| Percentage | 18.82% | 62.35% | 18.82% |

### *5.3 Discussion*

At the start of this design study, we formulated six design principles. In the coming paragraphs we look at the six design principles and assess to what extent these they had been met within the three cycles. At the end of this section Table 1.8 presents a summary of the different design cycles, the adaptations and what the adaptations yielded.

The first of these was that *the second sub-task must ensure that the students experience a setback, similar to setbacks they could experience at school under normal circumstances*. Two factors led us to conclude after the third design cycle that the untranslatable sentences indeed formed a setback for the students. First, the students performed the post-setback task significantly worse than the pre-setback task. In other words, there seemed to be an effect of the intervention. Secondly, in the classroom discussion all the students indicated that the second task was the most difficult. Moreover, no student had realized that it was an impossible task and everybody thought it was a normal task. We interpreted this as students finding the untranslatable sentences as a probable difficult task (i.e. setback) they might encounter at school (see also design principle 4: *All the subtasks must feel familiar to third year Latin students. This particularly holds for the second subtask)*. It seemed then that the untranslatable task could be used to simulate an academic setback.

The second design principle was: *the pre- and post-setback tasks must be internally consistent.* This was true for the pre-setback sentences of version B, but the internal consistency of version A was lower than .70. However, the scale was only three items and the sentences contain different grammatical features, so the internal consistency must be considered in this light (cf. Taber, 2018). According to Herman (2015) when an instrument is made up out of fewer than ten items the Cronbach's alpha underrates the internal consistency. Field (2017) has also discussed the relationship between more items and a higher Cronbach's alpha. Using a Spearman-Brown formula for testing whether the reliability would change if the number of items was increased, we found that the internal consistency would increase to .79 if the number of items was six instead of three (and .84 for the version B pre-setback sentences). To a certain extent the internal consistency of both sets of sentences could be deemed sufficient.

However, there was also a difference between the two sets regarding difficulty, which was not on par with the third design principle (*the pre- and post-setback task must be of comparable difficulty*). The sentences of the B version had proven to be significantly easier than those from version A. An adjustment must thus be made to one of the sets to make them more comparable. By adjusting the sentence from version A which affected the internal reliability the most negatively, both the difficulty and internal consistency problem might be solved. We thus sought to adapt the sentence *Equus pilo traiectus est*, before using the instrument again.

After re-examining the students' translation, it was decided that the problem arose from the word *traiectus.* This is a past participle form. This grammatical feature might be postponed in teaching at some schools, as there are no fixed requirements regarding when each grammatical feature is introduced in the Latin curriculum. Not being acquainted with this feature might have led students to not translating it sufficiently accurately. Indeed, after checking with the teachers, the Latin past particle had not been introduced to all the students from the study. An extra note in the wordlist informing the students to translate it as a past participle should solve this problem in the future. With this small alteration we trusted that the difficulty and internal consistency problems with the task should also be solved. This led us to also suppose that a next stage using the 'new' Latin Buoyancy Task with students thinking-aloud would be possible, particularly if the more difficult A-set is used as the pre-setback task.

The penultimate design principle was that *the first subtask must feel attainable to students, yet still be complex enough to encourage students to use (cognitive, metacognitive and affective) problem-solving skills and make thinking-aloud possible*. Seeing that most students in the third cycle said they found the pre-setback task the easiest, and the fact that, with the exception of the *traiectus* -sentence, the sentences scored >1 point it seems likely the task was attainable for them. Moreover, the sentences still presented the students with challenges as only few students were capable of translating all three sentences perfectly. Thus, we assumed that the Latin Buoyancy Task adhered to this design principle.

The final design principle was *the task must ensure that the participants are not hindered by a lack of vocabulary knowledge.* No more mention of the wordlists causing problems was made by the students after the first additions. In each cycle, the students did comment on the helpfulness of the wordlists. This led us to conclude that the final design principle was also met.

In Table 1.8 a summary is given of the different stages of the design process. The number of students testing the sentences, what was tested and how each task was altered compared to the previous cycle is the information included in the summary. The final version of the Latin Buoyancy Task itself, including the adapted wordlist can be found in Appendix 1.1.

*Table 1.8 Summary of the design process*

| Cycle | N | Tested | Design principles | Alterations made | Type of data collected |
|---|---|---|---|---|---|
| 1 | 28 | 12 translatable sentences | 4,5,6 | | Translations<br>Verbal remarks<br>Written notes |
| 2 | 26 | 6 translatable sentences | 4,5,6 | 2 grammatical & 1 wordlist alterations | Translations<br>Verbal remarks<br>Written notes |
| | 28 | 3 untranslatable sentences | 1,4 | | Translations<br>Verbal remarks<br>Written notes |
| 3 | 85<br>41<br>44 | All sentences<br>Version A<br>Version B | <br>1,2,3,4,5,6<br>1,2,3,4,5,6 | All three sets consecutively translated by students | Translations<br>Verbal remarks |

## 6 CONCLUSION

In this first part of this chapter, we presented the design study which resulted in the Latin Buoyancy Task. Up until now, research regarding academic buoyancy has relied on self-report or teacher driven measures. These methods however give rise to concerns about validity due to social desirability. Our aim was, therefore, to create a task-based instrument that triggered a setback and to further study academic buoyancy in gifted students based on data not based on self-report.

Simply put, the Latin Buoyancy Task consisted of three subtasks. These subtasks followed a translatable-untranslatable-translatable pattern to simulate a setback in the middle of the task. Six design principals formed the base for the task and we went through three different design cycles. A complete summary of the design cycles was presented in Table 1.8.

Primarily, none of the students seemed to have noticed that the untranslatable nature of the second task. All students did notice the difficulty increase. We, therefore, concluded that the task successfully mimics a setback. In the analysis of the final full design cycle, we found that the quality of the post-setback translations was significantly poorer compared to the pre-setback translations. Moreover, we found that the quality of some students' translations significantly decreased post-setback, whereas others seemed to have thrived. We also concluded that the A set of sentences was found to be more difficult than the B set. This coincided with a problem with the internal consistence within Set A. This led us to recommend adding an extra note explaining that *traiectus* is a past participle. This problem might be avoided to some extent by having all the participants perform the easier set B as the post-setback task.

By making these adjustments to the Latin Buoyancy Task, it should now be possible to use it as a task-based measurement in further investigations into academic buoyancy. The scores found in the third design cycle give us confidence that the translation tasks are attainable to third-year Latin students without being too easy. So, our next step with the task is to follow our recommendations and further use it in a think-aloud setting with students.

## 7 FURTHER DEVELOPMENTS

The data collected during the piloting of the thinking aloud (see Chapter 2) and the think aloud study itself, presented us with some insights into the effects of the final changes to the buoyancy task. We present the final changes and the gained insights in the rest of this section.

Interestingly, in the think-aloud study we did see a large improvement on the quality of the translations of the *equus* sentence: on average students scored 1,6 points ($N$ = 16) instead of 1.02 ($N$ = 41). We, therefore, assumed that the tasks were indeed made more comparable in difficulty by the addition of the *traiectus* note. Based on this, the A version was still a little more difficult than the B version, when comparing it to the results from the 44 students (who had an average score of 3,91 on the B version). As all the students in the main study made the tasks in the same order (with the slightly more difficult A functioning as the pre-test), this could not be checked for significance and remains only an observation.

# ANAGRAM TASK

## 1 INTRODUCTION

We aimed at administering the Latin Buoyancy Task under think-aloud conditions. For this, we were particularly interested in contrasting students who were affected by the setback task (i.e. non-buoyant students) with those who were not affected by it (i.e. buoyant students) to gain insights into the strategies of non-buoyant students. As thinking-aloud is a labour-intensive method of collecting data, the number of participants would be limited. Leaving the presence of both academically buoyant and non-buoyant students to chance in a small number of participants seemed imprudent. The aim of the selection was to particularly minimize students who were wrongly expected to be either non-buoyant or buoyant. To select a mix of buoyant and non-buoyant participants we thus needed a method of selection.

We determined that a method of selecting a mix of buoyant and non-buoyant students must be task-based, as the Latin Buoyancy Task was task-based. As academic buoyancy seems to be a general trait and not subject specific (Malmberg, Hall & Martin, 2013), we hypothesized that the students who were most affected by the selection task, would also be affected by the Latin Buoyancy Task. We therefore sought to design a linguistic task as a selection instrument, which was similar in set-up to the Latin Buoyancy Task, containing a pre-setback task, a setback and a post-setback task.

For this instrument we determined the following design principles:

1. Similar to the Latin Buoyancy Task, the selection task must ensure a setback;
2. Similar to the Latin Buoyancy Task, the internal consistency within the items of each sub-set must be sufficient;
3. Similar to the Latin Buoyancy Task, the pre- and post-setback task must be of comparable difficulty;
4. For students the task must not be too similar to the Latin Buoyancy Task or recognisably testing the same phenomenon;
5. The selection task must allow us to select participants for the think aloud study who would be particularly interesting regarding their extent of academic buoyancy.

Where the first three design principles are self-evident, the fourth and fifth need further explanation: if the Latin Buoyancy Task and the selection task were too similar, the students might work out that the middle part was manipulated. If that were to happen, the Latin Buoyancy Task would likely lose its intended effect. Regarding the fifth design principle, we were particularly interested in the students who were the most and least academically buoyant on the translation task, thus the task needed to allow us to filter out the middle group of students.

One of the ways in which the selection task differed from the Latin Buoyancy Task was the domain the task originated from. We did not choose a translation task, but another type of task that requires the employment of linguistic problem-solving skills: the solving of anagrams. Anagrams were chosen as unsolvable Anagram Tasks have previously been used in studies related to motivation. Lucas et al. (2015) isolated failure in a lab setting to study how grit affects students' perseverance. In one of their sub-studies they made use of a linguistic task in which they presented their participants with 37 anagrams, 16 of which were unsolvable. Earlier, using unsolvable anagrams Aspinwall & Richter (1999) concluded that optimists are able to lay aside an unsolvable task and focus their attention on a solvable one. Unsolvable Anagram Tasks have thus been shown to be a viable method in motivation studies, suggesting relevance in the case of academic buoyancy.

However, anagrams in the participants' own language are likely to be more familiar to students than Latin sentences. For, frustration level texts are familiar to Latin students (Boyd, 2017), making the impossible nature of the task less obvious. There was, thus, a risk of making them unsolvable if we aimed at using the anagrams as a selection task. Whilst trying to solve the unsolvable anagrams, the students might become aware that there were no solutions. If that was the case, they would not be affected by the setback task as intended. This would not only jeopardize the results of the selection task, but also the results from the Latin Buoyancy Task: the students would be aware that the Anagram Task was connected to the think aloud study, they might also make the transfer that if the second anagrams were impossible the second translation task might also be impossible. If students did make that inference, the effect of the setback would be negligible. Therefore, we did not include unsolvable anagrams, merely extremely difficult anagrams.

## 2 METHOD

In the following sections we present the development of the Anagram Task, which aimed at distinguishing participants that would be of particular interest for the think aloud study, regarding their academic buoyancy. The development of the task consisted of two design cycles, the first of which was mainly explorative. During that first cycle we asked 43 students to test all the anagrams and found that they were sufficiently solvable for third year students. We then used the results to create two comparable sets of seven solvable anagrams. In the following sections we present only the more extensive second cycle.

### *2.1 Participants*

125 students participated in the second design cycle. They were in their third year of pre-university level and their ages ranged from 14 to 16. The participants attended four different schools in the Netherlands. The students' parents were informed about the study and they were asked for passive consent prior to the data collection. All these students also participated in the testing of the Mindset Questionnaire and 85 of the students also participated in the third design cycle of the Latin Buoyancy Task (see pages 53 and 33 respectively).

### *2.2 Materials*

We compiled three sets of seven five-lettered Dutch anagrams using an online anagram tool. The pre-and post-setback tasks were similar in that all the anagrams included in those tasks had multiple solutions and a total of 25 solutions for each set of seven anagrams could be found. The second set was visually similar: seven clusters of letters. Of these seven clusters only three were actually solvable, and even then, there was only one solution per cluster, making a maximum score possible of three. The increase in difficulty was, therefore, large in the second task. The anagrams and their number of possible solutions can be found in Table 1.10.

*Table 1.10. The Anagram Test*

| Pre-setback task | | Setback task | | Post-setback task | |
|---|---|---|---|---|---|
| Letters | # Solutions | Letters | # Solutions | Letters | # Solutions |
| AEKRT | 3 | AGKLU | 0 | AEGLN | 5 |
| AETLR | 4 | BEFOR | 0 | AEGRV | 4 |
| EKLON | 3 | BEGLO | 1 | AEMNR | 3 |
| EKOPR | 4 | EIKLP | 0 | EEGIN | 3 |
| EMNOR | 3 | EKLNU | 1 | EKORT | 3 |
| EOPRT | 3 | ENNOX | 1 | EKOST | 4 |
| EOPST | 5 | HIPTU | 0 | ENORT | 3 |

### *2.3 Procedures*

The data was collected at the selected schools in a classroom setting. The students' teacher was present in an audience role. No significant incidents occurred during the data-collection. In the case of students participating in the testing of multiple instruments, the Anagram was given last. The students were first given a warm-up task containing four lettered anagrams. Further instruction included that the anagrams did not have to be solved in order.

The Anagram Task were randomly distributed to the students. Similar to the testing of the Latin Buoyancy Task, the pre- and post-setback tasks were inverted: around half the students made Set A as the pre-setback task, and Set B as the post-setback task, whereas this was inverted for the remaining students. The different versions were randomly distributed within the classes. This led to Set A being given to 62 students as the pre-setback task, and Set B to 63 students.

After all students had received the task, they were given two minutes to find as many solutions as they could per set. In the same debriefing session as that of the Latin Buoyancy Task, students were told about the increased difficulty of the second task. The students were asked during the post-classroom discussion what they felt was the connection between the Anagram Task and the Latin Buoyancy Task. Students answered for example that both tasked were based on language problem solving, or that translating Latin is a puzzle just like solving anagrams. None of the students mentioned that they were similar in that the second task was much more difficult until prompted by the researcher. For the students who did not also participate in the testing of the Latin Buoyancy Task there was also a debriefing session, which just focused on the Anagram Task. When asked which task was the most difficult the students all agreed on the second task. Some students felt the pre-setback task was easiest, whilst others thought the post-setback task was.

### *2.4 Analyses*

The students' solutions were scored using answer key the found in Appendix 1.2. Each correct solution was given a point, so there was a maximum of 25 points in the first and in the third part, and a maximum of three in the second part. To test for internal consistency among the items, only the first part made by the students was taken into consideration. Using SPSS v.25, an independent samples test was run to determine whether there was a difference between version A and version B.

To determine to what extent students bounced back after the second task, we first converted the scores into z-scores to make comparing between the two versions possible. We subtracted the scores from the post-setback task from those the pre-setback task. Students whose scores decreased more than a standard deviation compared to the pre-setback task were taken to be significantly affected by the setback, indicating that they are probably non-buoyant. Those whose scores increased more than a standard deviation were interpreted as being most likely to be buoyant. A paired-samples t-test was ran to determine if there was an effect of the second part and its significance, where Cohen's *d* was calculated for the effect size.

### *2.5 Results*

The internal consistency of the pre-setback task of version A was .70. On average, students found 5.21 ($SD$ = 2.67) solutions. For version B the internal consistency was .67. For this version, 5.17 ($SD$ = 2.29) solutions were found on average. In both sets, the students found most solutions for the first anagram. No significant difference was found between the scores of the two groups ($t$ (123) = .08, $p$ = .469). Moreover, for the setback task students found an average of .15 ($SD$ = .36) solutions.

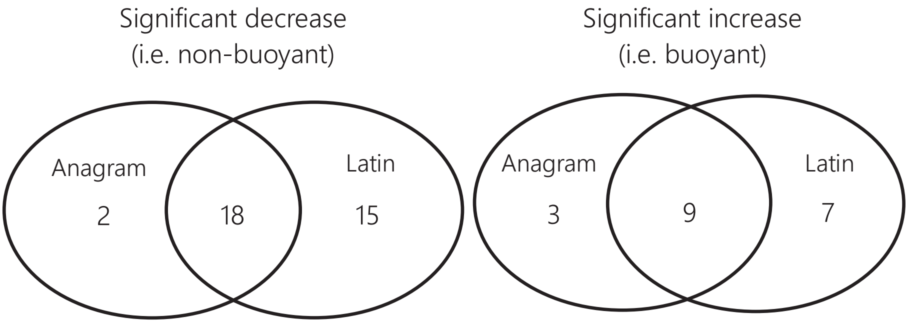
Regardless of version, the students found less solutions after the setback anagrams than before. We found that on average the version A students scored 1.61 points less ($SD$ = 2.62) and the B version students scored 1.05 points less ($SD$ = 2.73). A paired-samples t-test indicated that the difference between the first and third part of the task was significant, $t$ (124) = 4.66, $p$ = <.001. Using Cohens $d$, the effect size was found to be small ($d$ = .42). Running a repeated measures Anova we found that there was no significant difference between the two versions ($p$ = .17).

Of the 125 participants who completed the Anagram Task, 20 students (16%) scored significantly lower after the setback (Version A, at least 5 points decline; Version B, at least 4 points decline). When only considering the students who also tested the Latin Buoyancy Task, 18 students (21.18%) showed significant deterioration in both the Latin and the Anagram Task. A significant improvement means that the students were not affected by the setback. These students, therefore, were definitely academically buoyant in our definition. Nine students (10.59%) improved significantly on both tasks. Furthermore, there was a large group of students whose scores neither improved or decreased significantly.

According to the fift design principle, the Anagram task would have to enable us to select particularly interesting students for the think-aloud-experiment.

There was a weak correlation found between the difference scores from the Anagram Task and the Latin Buoyancy Task ($r$ = .28). However, the aim of the Anagram Task was not to be a *perfect* predictor of how students would perform on the Latin Buoyancy Task, but to filter out the middle group of students whose scores would not significantly change after the setback, thus providing us with the possibility to select participants on both extremes of the spectrum regarding academic buoyancy. We aimed at minimizing mispredictions of academic non-buoyant students and academic buoyant students. Figure 1.3 presents the overlap between the two instruments for the students whose scores significantly changed after the setback. We found that in the case of non-buoyant students (i.e. students who scored more than a standard deviation less on the post-setback task) only two students were wrongly expected to be non-buoyant according to the Latin task. For the students whose scores improved significantly on the Anagram Task, 25% did not improve significantly on the Latin Buoyancy Task. However, these students' scores did not significantly improve either.

*Figure 1.3 Overlap between non-buoyant and buoyant students according to the Anagram Task and the Latin Buoyancy Task.*




From this we concluded that the Anagram Task met the fifth design principle: it helps select participants who will probably be particularly interesting regarding their extent of academic buoyancy.

## 2.4 DISCUSSION

For this instrument we established five design principles. The first was that *the instrument would ensure a setback*. Seeing that the average score of the intended setback task was only .15 and this was much lower than the pre-setback task, regardless of version it we concluded that students noticed the large increase in difficulty. This might be taken as an indication that students indeed experienced a setback during the Anagram Task. Furthermore, the students all agreed in the classroom discussion that the second set of anagrams was the most difficult. We therefore assumed that the Anagram Task met this research principle.

The second design principle was that the internal consistency was sufficient. For version A this was only just the case (.70) and for version B, the internal consistency was lower than .70. When looking at the specifics of the task, the consistency measurements for the Anagram Task might have been affected since there were multiple solutions for each anagram and the students did not know how many there were. In both sets the students found the most solutions for the first item. Despite that the students were told that they did not have to complete the task in order, it seems that they focused mostly on the first items. This would explain the lower internal consistency.

Lucas, et al. (2015) found that some students even persisted on their anagrams to a fault and could waste time and effort on items instead of moving on to more fruitful pursuits. If students spent too long on the first task this could influence their finding solutions to the other items, hence affecting the internal consistency. If this is the case, it is not a solution to swap out the first items to improve the consistency. Seeing the consistency is not extremely far from the .70 despite the first item problem, the consistency might be considered sufficient given the nature of the task.

The third design principle was that the *pre- and post-setback tasks must be of comparable difficulty.* No significant difference was found between the pre-setback task scores of version A and version B. The instrument thus met this principle.

The penultimate design principle was that *the task is not too similar to the Latin Buoyancy Task or recognisably testing the same phenomenon for students.* Based on the classroom discussion, we believe that the students who performed both tasks on the same day did not see the overlap between the two tasks. Thus we assume the design principle was met. When using both instruments, however, it is still advisable to do so with some weeks between both tasks, to further ensure no effect of the first instrument on the second. Particularly as, if the Anagram Task is to be used as intended (i.e. to select with some certainty students who are non-buoyant), the Anagram Task is performed first by the students. It is possible that the students are more likely to notice the pattern of a particularly difficult second task. This could in turn affect the students' experience of the setback and their reaction to it on the think-aloud task.

The final design principle was the Anagram Task must allow us to select participants for the think aloud study who would be particularly interesting regarding their extent of academic buoyancy. Particularly, the task had to select participants that were the most and least academically buoyant and thus be of particular interest for the thinking aloud. The correlation between the two instruments was weak. However, of the 20 students who proved to be non-buoyant on the Anagram Task, only two did not significantly deteriorate on the Latin Buoyancy Task. The quality of both students' translation of the post-setback task did decrease above average, only not more than a standard deviation. In the case of significant improvement, there was a 75% overlap. The number of wrongly expected non-buoyant students based on the Anagram Task was thus minimal.

15 students did not perform the post-setback Anagram Task more than a standard deviation worse compared to their pre-setback task, but did do so on the Latin Buoyancy Task. These students might be non-buoyant students, who are missed when using the Anagram Task to select participants. This difference might be accounted for by the fact that anagram solving is not part of the Dutch curriculum, therefore, students do not regularly solve anagrams. For some students this might have led to an on-the-spot learning effect and as the task progressed they became more skilled at doing so. This is less likely the case during the Latin Buoyancy Task as they are already used to translating Latin regularly.

### *2.6 Conclusion*

In this part of the chapter we presented the development of the Anagram Task. The Anagram Task was to function as a selection tool for the think-aloud study. To do so, it aimed at particularly identifying which students were non-buoyant and which were buoyant on this task, thus selecting students for the Latin Buoyancy Task. To create the instrument, we formulated four design principles. Overall, the design principles of the instrument were met. Most importantly, the number of students wrongly thought to be non-buoyant based on the Anagram Task was small. Based on these findings the Anagram Task can perform its intended function of using it to select participants for the think-aloud study.

## 3 FURTHER DEVELOPMENTS

In the study we used the Anagram Task to select participants via extreme sampling. The findings from the design studies indicated that the Anagram Task was not a perfect predictor for how the participants would react to the setback on the Latin Buoyancy Task, with around 25% of the students performing differently. However, we decided that the overlap between the instruments was sufficient to justify using the Anagram Task as a selection instrument. After having conducted the think aloud study with the selected participants, it was possible to reflect on whether the Anagram Task indeed provided us with participants who were buoyant or non-buoyant as expected.

In Table 1.11 we present the findings from the study regarding the difference between the pre- and post-setback tasks on both the Latin Buoyancy Task and the Anagram Task. Four students, Bella, Claudia, Julia and Quintina, performed significiantly worse on the post-setback anagram task, but did not do so in the Latin Buoyancy Task. In the coming sections we take a closer look at each of these four students to determine whether they were rightly selected as participants.

*Table 1.11 Difference scores pre- and post-setback tasks, ranked from largest negative to largest positive change in Anagram Task*

| Pseudonym | Anagram Task | Latin Buoyancy Task |
|---|---|---|
| Julia | -6* | -1 |
| Anthony | -5* | -5* |
| Ennius | -5* | -5* |
| Octavia | -5* | -4* |
| Claudia | -4* | -1 |
| Quintina | -4* | +2* |
| Rufus | -4* | -4* |
| Bella | -3* | -2 |
| Diana | -3* | -3* |
| Gaius | -3* | -3* |
| Flavia | +2* | -1 |
| Horatia | +2* | -2 |
| Nona | +2* | -2 |
| Livia | +3* | +1* |
| Marcus | +4* | -2 |
| Phaedra | +6* | -2 |

*Note. * indicates a significant change, i.e. a change more than a standard deviation larger than the average change between the pre- and post-setback task.*

*Bella*

In the case of Bella, both her scores on the Anagram Task and Latin Buoyancy Task are borderline significant or not-significant. In both tasks, her scores deteriorated after the setback task, indicating that she was not particularly buoyant and thus a valid 'non-buoyant' participant for this study.

*Claudia & Julia*

Claudia and Julia's scores decreased significantly on the Anagram Task but did not do so on the Latin Buoyancy Task. For the Latin Buoyancy Task, they were neither particularly buoyant or non-buoyant. Claudia and Julia could be interpreted as mishits. In the qualitative Chapters 5 and 6 their scores did not cause to them being selected for further qualitative analysis of their translation process. Claudia's and Julia's translation processes were thus only minimally included in the study and therefore affect the results limitedly.

*Quintina*

The pattern found in Quintina's results of the Anagram Task and the Latin Buoyancy Task, is particularly striking: from a significant decrease in the Anagram Task, her scores significantly increased after the Latin setback. As we shall further divulge in Chapter 6, Quintina admitted to having adapted her usual task process whilst performing the Latin Buoyancy Task: due to the setting of the thinking aloud, she persisted longer than she normally would have, leading her to discover new ways of solving problems. These leads us to conclude that if she had performed the Latin Buoyancy in a similar setting as the Anagram Task, her scores would likewise have improved less or even not at all on the Latin Buoyancy Task. Therefore, the selection of Quintina as a non-buoyant participant was not redundant.

## MINDSET PREFERENCE

The next instrument we present in this chapter aimed at measuring mindset preference, so that we could include this preference as an individual difference between students in our analyses. Here we introduce the the development of a Dutch version of the Mindset Questionnaire.

### 1 INTRODUCTION

Beliefs about intelligence, affect how students learn. These beliefs about intelligence have been defined as 'mindset' and, according to Dweck, there are two mindsets (e.g. Dweck & Legget, 1988; Dweck, 1999; Dweck, 2011). On the one hand a student can hold a malleable view of intelligence. Such a student believes that intelligence is not something fixed, but something that can be developed. This way of thinking means the student holds a growth view of intelligence and therefore has a growth mindset. On the other hand, students can believe intelligence is static and that little can be done to change how intelligent they are. This so-called fixed style of thinking about intelligence leads students to have a fixed mindset. Having either a growth or fixed mindsets affects how students learn.

Mindset specifically plays a role in how students view and deal with difficult tasks. Studies have indicated that students with a growth mindset thrive when challenged: they perceive challenges as learning opportunities (Yeager & Dweck, 2012) and persevere (Duckworth, Peterson, Matthews & Kelly, 2007). By contrast, students with a fixed mindset are likely to give up quickly (Hong, et al., 1999) and find it important to be perceived as 'intelligent' (Dweck, 1999). Moreover, Moser, Schroder, Heeter, Moran & Lee (2011) demonstrated that students with a fixed mindset show less positive brain activity when they make mistakes than those with a growth mindset. These differences in how students view challenges led us to hypothesize that mindset preference might moderate the effect of the Latin Buoyancy Task on students.

Our targeted group of gifted students also made it relevant to include mindset in our study. Dweck (2012) implied that fixed thinking is something gifted students are particularly prone to. On the other hand, Ziegler & Stoeger (2010) have argued that fixed thinking only arises when gifted students relate failure to their own deficiencies. Similarly, Mofield & Parker Peters (2018) concluded that gifted students might not be more prone to fixed thinking and also found that gifted students had higher growth scores compared to non-gifted students. Despite the large body of literature relating to mindset, and mindset having received attention in the gifted field there is still much unclear how mindset functions in gifted students (c.f. McCoach & Flake 2018). We, therefore included mindset as a student characteristic in our study on academic buoyancy.

To include mindset as a student characteristic in our study, we needed a measurement to determine the students' mindset preference. A valid Dutch measurement would be the aim of this design study. Dweck and others have measured mindset preference using a self-report questionnaire, (e.g. Dweck, Chiu & Hong, 1995; Dweck, 1999; Romero, et al., 2014). Multiple versions of the questionnaire are in existence, and generally high levels of reliability have been demonstrated (e.g. Blackwell, Trzesniewski & Dweck, 2007; Boaler & Dweck, 2016). Translated versions of the mindset questionnaires have also previously been used successfully in research outside of English-speaking countries (e.g. Faria & Fontaine, 1997; Spinath & Stiensmeier-Pelster, 2001; Dupeyrat & Mariné, 2015; Abd-El-Fattah & Yates, 2006), but no Dutch instrument has previously been statistically tested. We therefore translated a mindset questionnaire and tested it in a Dutch setting.

We decided on determining mindset preference with a questionnaire, despite general concerns with self-report (Fan, et al., 2006) as we already had the Latin Buoyancy Task as a major task-based instrument for students to perform. Furthermore, we expected that the qualitative data from the thinking-aloud protocols could reinforce and help interpret the findings from the self-report. We thus expected to be able to overcome possible problems with self-report.

In the following sections we first focus on the first full design cycle. Then, another round of Method, Results and Discussion sections follows describing the second design cycle. An overview of the cycles, the changes to the questionnaire and the number of testing participants can be found per round in summarizing Table 1.17 at the end of the Design Cycles section (page 60). The final version of the questionnaire can be found in Appendix 1.3.

## 2 TWO DESIGN CYCLES

The development of the Dutch three-itemed Mindset Questionnaire had the form of a design study. Two cycles consisting of a (re)designing phase, a testing phase and analysis phase. In the following sections we focus on the first design cycle and then on the second cycle.

# DESIGN CYCLE 1

## *2.1 Method*

The aim of the first design cycle was to translate the Dweck mindset questionnaire into Dutch and test it for internal consistency.

### 2.1.1 Participants

125 cognitively gifted students participated in the first cycle of this study. These were the same students who participated in the testing of the Anagram Task (see Part II of this chapter), and 85 of these students also participated in the third design cycle of the Latin buoyancy test (see Part I of this chapter). They were all in their third year of pre-university level, ages ranging from 14 to 16 years. The participants attended three different schools in the Netherlands. The students' parents were informed about the study and asked for passive consent prior to the data collection. No refusal of participation was received from the parents.

### 2.1.2 Materials

The three-itemed questionnaire by Dweck, Chiu & Hong (1995) was selected as the source questionnaire. This questionnaire contains three statements that align with fixed believes regarding intelligence, of which participants are asked to what extent they agree with the statements. This version of the questionnaire seemed to be the most used measure (e.g. Blackwell, Trzesniewski & Dweck, 2007; Romero, et al. 2014; Yeager et al., 2019). Moreover, it has been argued that only using fixed items is sufficient when the goal is to divide students into mindset preference groups (Hong, et al., 1999). Finally, the short length of the questionnaire makes it particularly appealing for teachers to administer it during their lessons.

### 2.1.3 Procedures

The original questionnaire asks to what extent the responders agree with three fixed statements, using a six-point Likert scale ranging from 1, totally agree to 6, totally disagree. We translated the three items into Dutch by a bilingual translator. As an assessment, these Dutch statements were given to three other bilinguals, who translated the questionnaire back into English (c.f. Werner & Campbell, 1970; Brislin, 1986; Harkness, Villar & Edwards, 2010). This process was done twice to further ensure the correctness of the translation and led to a Dutch version of the questionnaire, that can be found in Table 1.12.

Our translations of the three statements deviated from the originals in one respect. 'You' (and its translation 'jij'/'je') can be interpreted as an impersonal pronoun or as a personal pronoun. The impersonal interpretation seems to be particularly strong in Dutch (De Hoop & Tarenskeen, 2015). The use of 'you' in questionnaire items, can lead participants to distance themselves from the question and interpret the item less through their own biases (Park, Ayduk & Kross, 2016). Using first person pronouns, on the other hand, strengthens the connection with self-concept and beliefs (Meissener, 2008). As we were actually interested in measuring the students biases towards intelligence and were also somewhat weary of receiving socially desirable answers, we decided on using the first personal pronoun. In doing so, we hoped to improve the reliability of the answers given by the students.

To test the Dutch Mindset Questionnaire, data was collected in a classroom setting by the researcher. The students' regular teacher was present in an audience role during the data-collecting and no significant incidents occurred during the data-collection. The questionnaire was distributed on paper and the students were asked to mark to what extent they agreed or disagreed with the statements. The students were given five minutes to fill in the questionnaire. At the end the students were asked to check if they had filled it in correctly in that 1 implied total agreement and 6 total disagreement. The students who also participated in the testing of the Latin Buoyancy Task, filled in the questionnaire prior to performing the Latin Buoyancy Task.

*Table 1.12 The original items and their Dutch translations*

| Original items | Dutch items |
|---|---|
| You have a certain amount of intelligence, and you can't really do much to change it. | Ik ben geboren met een bepaalde intelligentie en ik kan niet echt iets doen om dit te veranderen. |
| Your intelligence is something about you that you can't change very much. | Intelligentie is een onderdeel van mezelf waar ik weinig aan kan veranderen. |
| You can learn new things, but you can't really change your basic intelligence. | Ik kan nieuwe dingen leren, maar mijn basisintelligentie kan ik niet veranderen. |

### 2.1.4 Analyses

Internal consistency between the three items was rated. Per student, the mean score was calculated from the three items. The analysis of these scores followed that of Dweck's studies (e.g. Hong, et al., 1999): When the mean score was lower or equal to three, this implied the student had a fixed mindset preference. If, on the other hand, the mean score was a four or higher, the student was classed as having a growth mindset preference. Any score that fell between these boundaries was taken as an indication of the student not having a clear mindset preference.

## *2.2 Results*

The internal consistency among the three items of our Mindset Questionnaire was .68. The statements were all phrased as students with a fixed mindset would agree with, therefore, it was to be expected that the means of the three items would be comparable. However, this was not the case. As the means demonstrate in Table 1.13, the responses given to the third item seemed to deviate from those of the other two items. By deleting the third item, the internal consistency would slightly improve.

*Table 1.13 Scores from the Mindset Questionnaire (N = 125)*

| Item | *M* | *SD* | Cronbach's alpha item if deleted |
|---|---|---|---|
| 1 | 4.25 | 1.13 | .64 |
| 2 | 4.10 | 1.19 | .40 |
| 3 | 2.85 | 1.21 | .69 |

Using the means to establish the students' mindset preferences we found that:
36 students (28.8%) fell into the fixed mindset category;
61 students (48.8%) fell into the growth mindset category;
28 students (22.4%) fell into neither category.

## *2.3 Discussion*

The internal consistency of the items on the Mindset Questionnaire was lower than .70. For a three itemed questionnaire it is possible that the alpha is underestimated (e.g. Herman, 2015). Aron & Aron (1999) have even argued that specifically in psychology studies .60 could be adequate, whereas others have proposed that in social science studies such as education and in the case of exploratory studies a value of .60 is acceptable (Hair, Black, Babin, Anderson & Tatham, 2006). However, in a series of studies using the three itemed questionnaire Dweck demonstrated, the reliability of the questionnaire is high, variating between .94 and .98 (Dweck, Chiu & Hong, 1995). Moreover, in recent German research which aimed to (re)validate the reliability of the three-itemed Mindset Questionnaire in specifically adolescents, also found higher internal consistency than we did (.83), albeit lower than what was found by Dweck (Rammstedt, Grünning & Lechner, 2022). It therefore seems likely that the problem was connected to our translation.

This premise is strengthened by the fact that the mean of the third item was not similar to that of the first two items. This might be related to the translation or the Dutch context. In the original English questionnaire, the concept 'basic intelligence' is meant to mean the same as 'a certain amount of intelligence'. It is possible that the Dutch phrasings '*bepaalde hoeveelheid intelligentie'* and '*basisintelligentie'* are interpreted by the students as different concepts. It is, therefore, advisable to include the same phrasing in all three statements. This should improve the internal consistency of the questionnaire items sufficiently.

In sum, phrasing of the items in the Dutch Mindset Questionnaire in its current form needed to be adapted to improve the internal consistency. The current alpha was encouraging enough to assume that with adaptations the instrument can increase in reliability.

## DESIGN CYCLE 2

### *2.4 Method*

The development of the Dutch three-itemed Mindset Questionnaire had the form of a design study. Two cycles consisting of a (re)designing phase, a testing phase and analysis phase. In the following sections we focus on the second design cycle. The aim of this cycle was to improve the internal consistency of the items, making it more comparable to the original. When this second cycle was completed, the instrument seemed a valid enough measure to be included in the main academic buoyancy study.

#### 2.4.1 Participants

83 cognitively gifted students participated in the second cycle of this study. They were all in their third year of secondary school at pre-university level, ages ranging from 14 to 16 years. The participants attended four different schools in the Netherlands. These were the same schools that were involved in the first

cycle. The students' parents were informed about the study and asked for passive consent prior to the data collection. No refusal of participation was received from the parents.

### 2.4.2 Materials

Following our analysis of the first version of the Dutch translation of the Mindset Questionnaire, we adapted the first two items to also include the term 'basic'. The new statements, that were used in the second design cycle can be found in Table 1.14.

*Table 1.14 The new and old items of the questionnaire*

| Cycle 1 | Cycle 2 |
|---|---|
| Ik ben geboren met een bepaalde intelligentie en ik kan niet echt iets doen om dit te veranderen. | Ik ben geboren met een bepaalde basisintelligentie en ik kan niet echt iets doen om dit te verandere |
| Intelligentie is een onderdeel van mezelf waar ik weinig aan kan veranderen. | Intelligentie is een basisonderdeel van mezelf waar ik weinig aan kan veranderen. |
| Ik kan nieuwe dingen leren, maar mijn basisintelligentie kan ik niet veranderen. | Ik kan nieuwe dingen leren, maar mijn basisintelligentie kan ik niet veranderen. |

### 2.4.3 Procedures

To test this version of the Mindset Questionnaire, data was this time collected in a classroom setting by the students' tutor. This was done during a tutor lesson. The tutors were supplied with a letter including instructions for the data collection. In a consultation with the researcher before the data collection took place, they had gone through the letter together. The tutors were informed to place the students in test-setup and to read the instructions from the letter aloud. These instructions included telling the students they were testing an instrument, which would be used in a Latin translation study. They further included that the students would be testing the instrument by filling in a questionnaire and had five minutes to do so. At the end, the teachers asked the students to check if they had filled in the questionnaire correctly, with 1 implying total agreement with the statement and 6 total disagreement. After the data-collection the researcher checked in with the tutors, who stated that there no unusual incidents occurred during the data collection.

#### 2.4.4 Analyses

For measuring the internal consistency and determining the students' mindset preference based on the means, we followed the same steps as in the first cycle. To establish whether the changes to the phrasing of the statements also had effect on the outcome of the students' preferred mindset an independent t-test was run using SPSSv.25.

### *2.5 Results*

The internal consistency among the three items of the Mindset Questionnaire increased to .74. The statements were all phrased as students with a fixed mindset would agree with, therefore, it was to be expected that the means of the three items would be comparable. In the first design cycle, this was not the case. However, as demonstrated in Table 1.15, the means were now within .8 of each other, making them more similar. Table 1.15 also indicates that by deleting any of the items, the internal consistency within the scale would not improve.

*Table 1.15 Item scores from the Mindset Questionnaire (N = 83)*

| Item | *M* | *SD* | Cronbach's alpha if item deleted |
|---|---|---|---|
| 1 | 4.4 | 1.21 | .59 |
| 2 | 3.8 | 1.29 | .72 |
| 3 | 3.6 | 1.32 | .67 |

In Table 1.16 we present how many students absolutely and relatively reported which mindset preference. This is contrasted with the results from the first design cycle. Equal variance was assumed (.08) and the difference between the two groups was not significant ($t$ = (206) -1.27, $p$ = .21).

*Table 1.16 Reported mindset preference, contrasting cycles 1 and 2*

| | Fixed mindset | | No preference | | Growth mindset | |
|---|---|---|---|---|---|---|
| | # | % | # | % | # | % |
| Cycle 1 | 36 | 28.8 | 28 | 22.4 | 61 | 48.8 |
| Cycle 2 | 23 | 27,7 | 15 | 18.1 | 45 | 54.2 |

*Note. Cycle 1, N = 125; Cycle 2, N = 83*

## 2.6 Discussion

The altered phrasing had the desired effect, in that the internal consistency between the items improved to an acceptable level of .74. Despite this being acceptable, particularly for a three-itemed measure, this was still lower than the reliability found in the original (Dweck, Chiu, Hong, 1995). It was also lower than found in the recent German translation (Rammstedt, Grünning & Lechner, 2022). Of the three items, the first was now the least connected. The internal consistency might be further strengthened, by simplifying the translation to 'I have a certain amount of basic intelligence and I cannot really do anything to changes this'[1]. Seeing the current alpha, this was not strictly necessary, but might be advisable for future use to further develop the measure.

When looking at the number of students with either a fixed or growth mindset preference, we found a difference compared to the studies led by Dweck. In those studies, the number of high school students who reported a fixed mindset preference is similar to that of those with growth mindset (e.g. Boaler & Dweck, 2016; Blackwell, Trzesniewski & Dweck, 2007; Dweck, 1999). However, in our study, a particularly large group reported a growth mindset preference and a much smaller number of students reporting a fixed mindset preference. This was also the case in the first cycle. The independent t-test indicated that the means which determined mindset preference did not differ significantly compared to those of the participants of the first cycle. This seemed to imply that this was not caused by the instrument. However, the different distribution compared to Dweck's studies might be explained by social desirability (Duckworth & Yeager, 2015) or by the population of our participants in one of two ways.

[1] *Ik heb een bepaalde hoeveelheid basisintelligentie en ik kan niet echt iets doen om dit te veranderen.*

The first possible explanation was connected to the fact that our participants were cognitively gifted students. Our findings of relatively high numbers of students with a growth mindset were in line with other studies, that specifically focused on gifted students (c.f. Ziegler & Stoeger, 2018; Mofield & Parker Peters, 2018). However, Dweck's own studies indicate that gifted students are particularly prone to developing a fixed mindset (2012). A second explanation could be that our participants differed culturally from those in Dweck's studies (c.f. Tuleja, Beamer, Shum & Chan, 2011). Cultural differences are known to affect the generalizability of measurements (Taber, 2014). Also, in some of the existing mindset studies conducted in a non-English-speaking setting results have been found that are not in alignment with Dweck's results (e.g. Faria & Fontaine, 1997; Spinath & Stiensmeier-Pelster, 2001; Dupeyrat & Mariné, 2005; Abd-El-Fattah & Yates, 2006). Thus, the giftedness or culture of our participants might be the key to explaining the large group of students reporting a growth mindset. That is why we interviewed 16 participants as to why they had filled in the questionnaire as they had (c.f. Quihuis, Bempechat, Jimenez, & Boulay, 2002).

*Table 1.17 Summary of the design cycles of the Mindset Questionnaire*

| | Students | Tested | Alterations to previous cycle |
|---|---|---|---|
| Cycle 1 | 125 | 3 statements | |
| Cycle 2 | 83 | 3 statements, 2 of which were adapted | Included the concept of 'basisintellegentie' in all the statements. |
| End version | | | Suggestion: simplify the first statement. Interviews |

## *2.7 Interviews*

In the previous Discussion section, we concluded that the Dutch mindset instrument was usable in its current form, but that it might benefit from an adjustment regarding the first item. We also announced that at a later date we had some of the students on why they answered the questionnaire as they did, to gain more insights into why it yields such a large group of students with a growth mindset.

We made use of the 16 students who participated in the think-aloud study to conduct brief interviews about the Dweck questionnaire. These were the students who participated in the second design cycle. During the interviews we asked them why they believed that intelligence was malleable or fixed according to what they had answered on the questionnaire.

Despite only having questioned 16 students, there was a clear common thread in their explanations. Students seemed to associate 'intelligence' with the level of education followed. In the Netherlands, once students leave primary education (usually at the age of 11 or 12), they follow all their classes at a certain level. These levels include VWO (pre-university education, 6 years), HAVO, (pre-college education, 5 years) and VMBO (pre-vocational education, 4 years). At the end of the year it is usually possible to move up or down a level, depending on the students' results. Furthermore, when students have passed their final exams, they can extend their high-school education by two years and move on to the next level. During the interview, all the students who reported a growth mindset gave examples of friends who had moved up or down a level when explaining why they thought intelligence is malleable.

It thus seems that the Dutch education-system might play a role in how the students (mis)interpret the mindset items and their understanding of intelligence. This means that it is probable that the difference in the measurement's outcome is indeed, at least in part, caused by a cultural difference between the participants and Dweck's results cannot directly be compared to Dutch studies regarding mindset.

### *2.8 Conclusion*

This second part of the chapter focused on the design of a Dutch version of the three-itemed Mindset Questionnaire. Table 1.17 presented a summary of the different design cycles. We hypothesized that mindset preference might moderate the effect of the untranslatable task. However, until now, there was no statically tested Dutch measurement for mindset preference. Therefore, we aimed at developing a statistically sound Dutch translation of the often used three-itemed Mindset Questionnaire.

In the second version the internal consistency was acceptable. A possible further improvement might be made to the internal consistency by simplifying the translation of the first item. Despite possible further improvements to the instrument, the current analysis indicates that the questionnaire can be used to determine mindset preference. In the think-aloud study we therefore use the questionnaire in its current form. In the study we use the results from the Mindset Questionnaire to relatively assort the students according to the extent of their mindset preference.

A noteworthy observation in the results was that a particularly large group of students reported a growth mindset preference. This outcome differs from Dweck's findings but does align with other studies that focused on gifted participants. To explore whether and to which extent other factors than giftedness might explain the disparate distribution of mindset preference in Dutch students we conducted preliminary interviews regarding the questionnaire. From these interviews it seems increasingly likely that students are affected by cultural differences and results from translations of the Mindset Questionnaire cannot be compared one-on-one to Dweck's findings. Further interviews about mindset and the questionnaire might further strengthen this finding.

# 1 INTRODUCTION

During the preparations of the think-aloud study, one unanticipated detail noticed whilst observing the students perform the Latin Buoyancy Task during the third design cycle, remained predominantly in our minds: in all classes there were students who showed physical signs of being affected in some manner by the setback task, particularly the fidgeting and sighing that was seen during the untranslatable task, but not during the pre- or post-setback task. Similar behaviour was then also seen in some students during the practice round of letting students perform the Latin Buoyancy Task whilst thinking-aloud, prior to the data-collection.

The behaviour seemed to imply a feeling of frustration and/or disengagement from the task in these students (Farley, Risko & Kingstone, 2013). This seemed a fitting interpretation of the fidgeting at that moment as frustration is an emotion that people can experience when they feel they do not have control over hurdles that impede them from achieving something (Anderson & Bushman, 2002; Pekrun, 2006). Observing the fidgeting led us to wonder whether we should include frustration tolerance as a learner variable in the study, besides mindset preference, as an indication of individual difference between students.

Students often experience frustration during academic tasks (Jonassen & Grabowski, 2012; D'Mello, 2013). Low frustration tolerance can lead to students avoiding challenges or procrastinating (Solomon & Rothblum, 1984; Bridges & Roig, 1997; Harrington, 2005; Walonoski & Heffernan, 2006; Wright, Lam & Brown, 2009). They might also give up when challenged, (Hoza, Waschbusch, Owens, Pelham & Kipp, 2001). Moreover, low levels of frustration tolerance are correlated with having a fixed mindset (Meindle et al., 2019), one of our included student characteristics.

We, therefore, decided to include frustration tolerance in our student characteristics at a later stage in the process than the other characteristics. This meant that we would design a measurement to determine the students' level of frustration tolerance, and that the measurement could only be tested minimally before being used for the data-collection. It was not ideal, but, the source instrument already existed and had been statistically validated. Thus, we embarked on the development of a measurement to determine the extent of frustration tolerance in students. In the end, the validity of the instrument left us with little concern: the internal consistency of the instrument was high, when compared to data from the think aloud study it indeed seemed to measure frustration tolerance (see Chapter 4) and no new questions arose regarding the instrument itself. One design cycle was thus sufficient for this instrument. In the following sections we present the design of the Frustration Tolerance Questionnaire. The final version of the questionnaire can be found in Appendix 1.3 alongside the Dweck Mindset questionnaire items.

## 2 METHOD

As the other learner variable included to account for individual differences (mindset preference) was measured by a questionnaire, and at the time, a general lack of task-based frustration tolerance measures that were useable in an educational setting, we decided to determine frustration tolerance on the basis of a questionnaire.[1]

### *2.1 Participants*

83 students participated in the selection procedure, after their parents had given passive consent. 81 of those students filled in the questionnaire completely, whereas two did not, for unknown reasons. The data we have on the Frustration Tolerance Questionnaire emanates therefore from 81 students.

### *2.2 Materials*

We based the questionnaire on one used by Wright, Lam and Brown (2009). Their items aimed at establishing whether their participants had a high or low frustration tolerance, specifically in situations where effort is put in without rewards ($\alpha$ = .72). They found that motivation decreased after non-reward when frustration tolerance was low. We chose this scale as we believed that the untranslatable sentences were an example of a task which necessitated effort, that was not rewarded.

We translated the items from the questionnaire to Dutch. These items were translated back to English by other bilinguals (c.f. Werner & Campbell, 1970; Brislin, 1986; Harkness, Villar & Edwards, 2010). We then adapted the items to specifically represent school situations, thus making the items recognisable to students. These items were statements that the students had to rate to what

[1] *In the meantime, a task-based instrument for an educational setting has been designed and validated, c.f. Meindle, et al. 2019.*

extent they agreed with them using a six-point Likert scale, ranging from 1 totally agree to 6 totally disagree. The eight items as we adapted them to school settings and the Dutch translations can be found in Table 1.18.

*Table 1.18. The adapted frustration tolerance items and the Dutch translation*

| Dutch | English |
|---|---|
| Als ik door omstandigheden een belangrijk doel niet kan halen, vind ik het moeilijk om het toch te blijven proberen. | If because of circumstances I cannot reach a certain goal, I find it difficult to keep trying. |
| Als ik een uitleg niet snap, haak ik af en let ik niet meer op. | If I do not understand an instruction, I turn of and pay no more attention. |
| Als iets waar ik naar uit kijk niet doorgaat, dan heb ik geen zin of energie om een alternatief te regelen. | If something I am looking forward to is cancelled, I cannot be bothered to arrange an alternative. |
| Als ik een toets heel moeilijk vind, beantwoord ik de vragen maar gewoon niet meer. | If I find exam questions difficult, I just don't answer them. |
| Als het niet gaat zoals ik wil, kunnen mijn dagelijkse bezigheden me minder schelen. | If something goes differently to how I want it to, I care less about my daily routine. |
| Als ik ergens hard aan heb gewerkt, maar niet de waardering krijg die ik zou verdienen, dan ben ik minder gemotiveerd. | If I worked hard for something, but I did not receive the praise I deserve, I lose my motivation. |
| Als iets leuks onverwacht toch niet doorgaat, ben ik een tijd lang minder opgewekt. | If something fun is unexpectedly cancelled, I am less cheerful for a while. |
| Als ik merk dat mijn huiswerk moeilijker is dan ik verwachtte, word ik chagrijnig. | If my homework is more difficult than I expected, I become grumpy. |

## *2.3 Procedures*

The instrument was implemented at the same time as the second cycle of the Mindset Questionnaire. Thus, the procedures agree with those presented in the Procedures section of the Mindset Questionnaire sub-chapter (page 57).

### *2.4 Analyses*

We combined the items from the Frustration Tolerance Questionnaire with the three items from the Mindset Questionnaire, making one questionnaire to test two different variables. The answers to the frustration tolerance items were combined again for analysis. Internal consistency was checked using SSPS v.25.

## 3 RESULTS

The internal consistency of our items was .83. This is a good consistency, while it also indicates that there are no redundant items (Taber, 2018). In Table 1.19 the internal consistency of responses can be found per item.

Overall the mean was 3.56 (*SD* = .87). On the one hand, 25 of the students' mean were higher than 4.0, leaning them more towards being highly tolerant of frustration. On the other, 21 students reported a mean of less than 3.0. They thus leant more towards a low tolerance. We also ran a paired t-test, comparing the means from the mindset items with the frustration tolerance means. They were not correlated, $r(80) = -.31$ $p = .38$.

*Table 1.19 Contribution to internal consistency per item (N = 81)*

| | Frustration tolerance items | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| M | 3.58 | 4.10 | 3.74 | 4.75 | 3.84 | 2.62 | 2.80 | 3.05 |
| SD | 1.17 | 1.36 | 1.37 | 1.47 | 1.27 | 1.08 | 1.05 | 1.40 |
| Alpha | .84 | .80 | .80 | .79 | .78 | .82 | .81 | .81 |

## 4 DISCUSSION

According to Wright, Lam and Brown (2009), motivation decreases in students with low frustration tolerance after putting effort into a task, without receiving a reward. Our first findings also seemed to point towards an interaction effect between the extent of frustration tolerance in a student and their motivation after the setback task (for more, see Chapter 4). This seemed to imply that our translated and adapted version of the Frustration Tolerance Questionnaire measured frustration tolerance.

The lack of correlation between the Dweck items and the frustration tolerance items was somewhat unexpected, seeing the two constructs are related

(Meindle, et al. 2019). However, our results from the mindset items differed from those from American-based studies in that the reported growth mindset group was much larger. Therefore, a lower correlation was not strange. It did confirm that mindset and frustration tolerance are separate self-regulation constructs, though related to each other in some way.

The Frustration Tolerance Questionnaire, thus, seemed promising as a measure. The current results do not give rise for suggestions on improving the questionnaire. However, this is only based on one testing and a relatively small group of participants ($N$ = 81). The instrument would benefit from further validation. Until then, the results must be interpreted with discretion.

## 5 CONCLUSION

Based on our first findings, the Frustration Tolerance Questionnaire seemed promising as a measure. However, not having further tested it, but only having used it once, we must interpret and use the data collected with this instrument with some caution.

# AFFECT

## 1 INTRODUCTION

During the preparations of the think-aloud study, and particularly during the testing of the Latin Buoyancy Task, we came to ask whether the current instruments would sufficiently yield data on the students' affection. The aim of the study was to establish how a temporary academic setback affects gifted students' use of cognitive, metacognitive and affective strategies. By having the participants perform the Latin Buoyancy Task while they were thinking aloud, we supposed that the students would particularly demonstrate their cognitive and metacognitive strategies. According to Van de Velde, van Keer, Schellings & Van Hout-Wolters (2015) the use of affective judgements is part of the reflection stage (c.f. Zimmerman, 2000), but not all students have mastered automatic self-regulatory learning. For our study, this meant that not all students would mention how the Latin Buoyancy Task made them feel without being prompted to do so. This then would leave the affective strategies underexposed in our data-collection.

However, affective strategies were of particular interest to our study as the Latin Buoyancy Task was aimed at triggering a setback. We expected experiencing a setback to be paired with emotions. Moreover, the role that affective strategies play during learning is gaining more attention from the educational research (e.g. MacCann, et al., 2020, Coughlan, Lister & Lucassen, 2021). According to Veine et al. (2020) the emotions caused by situations that might be described as a setback 'demand a reflective approach' (p. 148). Furthermore, these emotions affect the learning process and task outcome (Pekrun & Linnenbrink-Garcia, 2012). Using the four instruments described in Part I-IV, we were at risk of not explicitly collecting data pertaining to the students' affective strategies. Thus, we had to strengthen our overall research design by including a method of collecting this data.

During the various design cycles, we had included classroom discussions. During these, the students shared information on how the tasks made them feel and how well they felt it went. These discussions not only provided information about how to improve the instrument, but also how the task had an effect on the students affectively. Thus, to still be able to gather information with the Latin Buoyancy Task on the affective state of the students, we decided to simulate the classroom discussion by adding four prompted reflection moments in the task. The first of these prompts was before the students started translating the pre-setback task, and the remaining prompts were made when each of the three translation tasks were completed.

We chose emoji as a representation of emotions to measure the students' affective state, because they require little explanation to students. Moreover, emoji's give teenagers help and autonomy in verbalizing their emotions (Mackenzie, Macdougall, Fane & Gibbs, 2018; c.f. Schouteten, Verwaeren, Lagast, Gellynk & De Steur, 2018). We hence created an emoji chart, containing 20 different emoji which represented emotions and moods the students might feel throughout the different stages of the task. The chart contained positive, neutral and negative emotions/moods (Pekrun & Linnenbrink-Garcia, 2012). The addition of the emoji chart in our study should guarantee at least some insight into how the students were feeling in addition to the think-aloud protocols.

According to Miller et al. (2016) emoji can be interpreted differently by different people. However, by including asking why the participant chose the emoji in the retrospective thinking aloud, we trusted that the student's intention with the emoji should be clear. Wilkinson, Carter, Satchwell and Bray (2021) also nuanced their findings by including further enquiry to why their participants had chosen an emoji. They also concluded that the emoji seemed to set their participants at ease. This observation is valuable to our study, for, thinking aloud necessitates that the participants feel at ease (e.g. Ericsson & Simon, 1980). Moreover, emotions tend to be more personal and thus difficult to share compared to other thoughts (Pickard, 2003). By making the participants feel more at ease, the Emoji Chart might also help overcome the personal nature of emotions.

We hypothesised that at the onset of the Latin Buoyancy Task our participants would have little reason to report negative emotions. However, as the task progressed, particularly after the setback task, we did expect an increase in negativity. By following the reported emoji, we hoped to gage the effect of the Latin Buoyancy Task on the participants emotive state and it thus support our understanding of how the participants' affective strategies were affected. In Chapters 4-6, we do not analyse the reported emoji statistically, but in Chapters 5 and 6 we include the emoji to support the qualitative interpretation of other data.

Similar to the Frustration Tolerance Questionnaire, the Emoji Chart was only administered during the data-collection for the main study. In the following sections we further explain and discuss the creation, administering and the gathered results from the Emoji Chart.

# 2 METHOD

The Emoji Chart was developed in the same stage of the study as the Frustration Tolerance Questionnaire. Thus, it was incorporated later in the study's design than the Latin Buoyancy task and therefore did not go through multiple design cycles. However, we did incorporate it in a trial run of the thinking aloud (see Chapter 2, p. 89). The feedback we received was that it was sometimes difficult to only pick one emoji. We incorporated this feedback by focussing less on the word 'one' in the task description and if necessary, allowing students to fill in multiple emoji. For more on this, see Chapter 2. As we shall see, this was not necessary in the case of the thinking aloud proper. In the following sections we focus on the data collected in the think aloud study and not the trial run, as that only included one participant.

## *2.1 Participants*

16 participants participated in the think aloud study and thus filled in the Emoji Chart. The participants were all third-year Latin students. Active consent for participation in the study had been given by their parents. We had purposefully selected the participants from a larger group of 83 students based on their results from the Anagram Task. For more on the participants and how they were selected we refer to Chapter 2.

## *2.2 Materials*

To create the Emoji Chart, we selected 20 emoji from among the most used on *WhatsApp*. We only included emoji that were based on faces, as we felt these best represented emotions. We based our selection on emotions that might occur in a school setting before and particularly during performing the Latin Buoyancy Task. Furthermore, we strived for a balanced mix of positive, negative and neutral emotions, so not to steer the results. We, thus, selected eight positive, eight negative and two neutral emoji.

The emoji and the emotions they represented are presented in Figure 1.4. In using the chart during the data-collection, we only included the emoji themselves and not the verbal reference to what they aimed to represent.

*Figure 1.4 Emoji included in the Emoji chart and the emotion they represented*

| Emoji | Emotion | + - o* | Emoji | Emotion | + - o* |
|---|---|---|---|---|---|
| 😡 | Angry | - | 😶 | Shy | o |
| 🤯 | Confused | - | 😎 | Confident | + |
| ☹️ | Disappointed | - | 😊 | Content | + |
| 😤 | Frustrated | - | 🤔 | Curious | + |
| 😒 | Irritated | - | 🤨 | Determined | + |
| 😴 | Sleepy | - | 🤓 | Focussed | + |
| 😩 | Unmotivated | - | 😃 | Happy | + |
| 😟 | Worried | - | 😌 | Hopeful | + |
| 😕 | Unsure | - | 🧐 | Interested | + |
| 😐 | Neutral | o | 😛 | Playful | + |

*Note.* * + stands voor positive, - stands for negative and o stands for neutral.

### *2.3 Procedures*

During the think aloud data collection, we asked the students to fill in the Emoji Chart four times. The first time was after the researcher had introduced herself and had explained that the participant would be going to translate nine Latin sentences, whilst thinking aloud, but before the think aloud warm-up task. The participants were asked to place a '1' next to the emoji that they felt best represented how they were feeling at that moment. We also asked the participants to mark a '2' after completing the pre-setback task, a '3' after the setback task and a '4' after the post-setback task. For more on the procedures of the data-collection, see Chapter 2.

Rufus said that he felt unable to choose an emoji when asked to do so at the beginning. When given the option to mark more than one emoji the participant still did not fill in the chart. We did not push the participant to do so further. The participant was able to choose an emoji during the later prompts. No other incidents occurred during the data-collection.

## 3 RESULTS

In Table 1.20 we present the emojis that the participants marked as best-fitting their mood at that moment. Looking at the emoji as the 16 participants filled them in throughout the different stages of the Latin Buoyancy Task, we see a pattern in how the emoji change. Particularly after the setback task, the emoji are nearly always negative. Only Diana and Marcus reported feeling 'curious' and 'interested' respectively after the setback task. Notably, after the post-setback task, all participants reported more positive emoji again. This is in accordance to our expectation that the setback task would affect the participants' emotions and, thus, their choice for a particiluar emoji.

*Table 1.20 The reported emoji*

| Student | At the start | | After the pre-setback task | | After the setback task | | After the post-setback task | |
|---|---|---|---|---|---|---|---|---|
| Anthony | 😃 | + | 😕 | - | 😤 | - | 😌 | + |
| Bella | 🤨 | + | 😊 | + | 😶 | o | 🤔 | + |
| Claudia | 🤓 | + | 😕 | - | 🤯 | - | 🤨 | + |
| Diana | 😴 | - | 🧐 | + | 🤔 | + | 😃 | + |
| Ennius | 😃 | + | 😃 | + | 😒 | - | 😃 | + |
| Flavia | 😛 | + | 🤯 | - | 😕 | - | 😎 | + |
| Gaius | 😃 | + | 🧐 | + | 😤 | - | 🤨 | + |
| Horatia | 😊 | + | 😕 | - | 🤯 | - | 😎 | + |
| Julia | 🤔 | + | 😕 | - | 😤 | - | 🤯 | - |
| Livia | 🤓 | + | 😕 | - | 😕 | - | 😊 | + |
| Marcus | 🤯 | - | 🤔 | + | 🧐 | + | 😃 | + |
| Nona | 😃 | + | 🤯 | - | 😕 | - | 😊 | + |
| Octavia | 😩 | - | 🤯 | - | 😤 | - | 🤯 | - |
| Phaedra | 😎 | + | 🤔 | + | 😕 | - | 😎 | + |
| Quintina | 😴 | - | 😕 | - | 😕 | - | 🤔 | + |
| Rufus | | | 🤔 | + | 😕 | - | 🤔 | + |

## 4 DISCUSSION

As emoji are subject to personal interpretation (Miller, et al., 2016), classifying them by adding a word to describe the emotion, is slightly tricky. As we did not want to influence the participants in their expressing what they were feeling by adding words to the Emoji Chart, we were dependent on the emoji themselves. By including the Emoji Chart in the retrospective thinking, we were more aware of the students' interpretation of the emoji and the participants often described the emoji verbally. For example, Anthony summarized that he felt 'hopeful'[1] after

[1] *Hoopvol*

the last task, this agreed with the word that we had labelled that particular emoji as for ourselves.

However, our aim with the Emoji Chart was not to exactly specify in words what the participant was feeling. Its purpose was to gage how the students' affective strategies were influenced by the Latin Buoyancy Task. The emoji on the chart could be sorted into positive, negative or neutral emotions. By doing so, the course of the participants' affective state could be roughly followed throughout the task independent of the exact emotion the emoji represented.

Rufus did not report the best fitting emoji at the start of the task, but had no problems doing so later in the task. This might be an example of the participant not yet feeling fully at ease at the beginning of the task and therefore being less willing to share how they were feeling (Pickard, 2003). According to Peterson (2015; 2018) gifted students are often unused to feeling that they are understood by others, and building trust necessitates more time than in the case of non-gifted students (Winsor & Mueller, 2020). Possibly not knowing the researcher or feeling apprehension towards the study affected student 16 at first to not wanting to share how he was feeling. Possibly other students also felt a little uncomfortable at first but did fill in the chart due to gifted studies having a high sense of obligation (Winsor & Mueller, 2020). However, as we only had one clear instance of a participant not filling in the Emoji Chart, and as the general course of the emoji fitted the Latin Buoyancy Task's pattern, we assumed that the answers given by the students were generally reliable.

We only used the Emoji Chart minimally. Further validation of the Emoji Chart might be done by organizing interviews with students to ascertain what emotion the emoji represents for them. A study similar to that of Wilkinson, et al. (2021) which included comparing and contrasting how different groups understood the same emoji would also be possible. For the current dissertation, we use and interpret the data from the Emoji Chart sparingly.

Looking at the results as presented in Table 1.20, the Emoji Chart does provide an indication of the participants' mood. Overall, the participants' moods were positive at the onset of the task. In the few cases in which the mood was negative before the task, this provided an indication of the general state of the participants' mood, which might affect how they engage with the task at hand.

## 5 CONCLUSION

During the preparations of the think aloud study we concluded that our instruments might not yet be sufficient to fully establish how the participants affective strategies were affected by the setback. This led us to develop the Emoji

Chart. With the Emoji Chart we aimed at gaging how the participants' affective strategies were influenced by the Latin Buoyancy Task. By not focussing on the exact words for the emotions, but by approaching them as 'more positive, more negative or more neutral emotions', we could follow the development of the participants' affective state throughout the Latin Buoyancy Task. In the course of this dissertation the Emoji Chart data is used with care, due to the personal interpretative nature of emoji.

## SUMMARY

In this chapter we presented the design of five different instruments: the Latin Buoyancy Task, the Anagram Task, the Mindset Questionnaire, the Frustration Tolerance Questionnaire and the Emoji Chart. These instruments were central to the data-collection for this dissertation.

The Latin Buoyancy Task was found to provide students with a setback halfway through the task, making it possible to isolate the setback and study the process of academic buoyancy using a task-based instrument. The Latin Buoyancy Task was performed by students whilst thinking aloud to gain further insights into academic buoyancy. The outcomes of the think-aloud study are presented in Chapters 4, 5 and 6.

To ensure that we included participants who were non-buoyant in the think-aloud study, we designed a task that would serve as a selection instrument. The Anagram Task proved to fulfil the aim of minimizing students who were wrongly expected to be non-buoyant. The Anagram Task, therefore, functioned as the selection instrument.

To determine whether the student characteristics moderate the effect of the Latin Buoyancy Task we also translated and adapted two questionnaires to ascertain the students' mindset preference and their extent of frustration tolerance. Both questionnaires proved to be useable in their current form to divide students into groups regarding their mindset preference (growth versus fixed) and frustration tolerance (high versus low).

Finally, we presented the Emoji Chart. The aim of this instrument was to supplement our data regarding the affective strategies the participants used throughout the Latin Buoyancy task. With the Emoji Chart complete, the think aloud study was fully prepared to attempt to answer to what extent the participants' cognitive, metacognitive and affective strategies were influenced by the setback task and how this is moderated by mindset and frustration tolerance.

# CHAPTER 2

# TAPPING INTO STUDENTS' TRANSLATION PROCESS I: THE DATA COLLECTION

The focus of this dissertation is to establish how a lack of academic buoyancy in gifted students is reflected in their translation processes. Academic buoyancy thus lays at the heart of this study. In the previous chapter we presented the five instruments which were implemented to gather data. The current chapter deals with how the study's data was collected using those instruments in a mixed-methods design. Self-report questionnaires were combined with data from concurrent and retrospective thinking aloud. In this chapter we explain why and how these methods were used. More importantly, we discuss how the findings from these different methods related to each other. Particular attention is given to the think aloud process, as this formed our main method of collecting data on academic buoyancy. We also include how we made use of extreme sampling to select the 16 gifted participants for the think aloud process. We explain our choice for gifted students as participants, as well as an operationalising of the complex term 'giftedness'.

# 1 INTRODUCTION

The focus of this dissertation is to establish how a lack of academic buoyancy in gifted students is reflected in their task process. In Chapter 1 we presented the instruments we designed to answer our questions relating to academic buoyancy. The aim of the current chapter is to outline how these instruments were implemented to gather the data needed for the study. Particular attention is given to the conducting of the Latin Buoyancy Task, with which data was gathered using a think aloud process (sometimes also referred to as TAP). The second, related, aim of this chapter is to describe the screening process of the participants including explaining how we operationalised giftedness.

Having participants think aloud whilst performing a task has long been a popular research method in psychological (e.g. Malek, Berna & D'Argembeau, 2017; Güss, 2018; Leighton, 2021) and educational research (e.g. van den Bergh & Rijlaarsdam, 2001; Kesler, Tinio & Nolan, 2016; Kim & Bowles, 2019; Rogiers, Merchie, Van Keer, 2020). Two factors are key for researchers when making the decision to generate data by thinking aloud. First, the aim of the study must be to assess the actual behaviour of the participants. Secondly, think aloud data are seen as valuable in particularly the explorative stages of studying that specific behaviour (Cf. Ericsson & Simon, 1993; Young, 2005). These factors both apply to our study: for reasons explained in Chapter 1, we did not want to study academic buoyancy using self-report, but by investigating actual behaviour. Moreover, research into academic buoyancy is still relatively sparse, and, to our knowledge, not yet investigated in the Netherlands. This holds especially true for a Latin translation task. This study was thus of an explorative nature. Therefore, making use of a think aloud method to collect data for our study seemed particularly fitting.

This rationale was further strengthened by the fact that we were interested in the effect of a difficult task on the participants cognitive and metacognitive processes. According to Ericsson & Simon (1980; 1993), the major proponents of this methodology, thinking aloud makes it possible for researchers to gain information on the participants' short-term memory. Thus, it can provide an insight into specifically the cognitive processes that are utilized when performing a task. Also, for studying metacognitive processes the importance and value of thinking aloud methods has often been emphasized by Veenman (e.g. 2011; c.f. Veenman, Elshout & Groen, 1993; Veenman, Prins & Verweij, 2003; Al Qahtani, 2020). This point of view was supported by the findings of a metacognition literature review study by Craig, Hale, Grainger and Stewart (2020). They analysed multiple studies and confirmed that students are often not cognizant of certain cognitive processes or unaware of automated processes. Although students are often unable to report such behaviour themselves, they can however demonstrate it. We, therefore, did not just opt to implement a think aloud method to collect part of our data, but to implement it for the data at the heart of the study.

A final argument for adopting a think aloud method in this study is the instrument itself. For, the Latin Buoyancy Task loans itself particularly well for thinking aloud. According to Payne (1994), tasks that contain verbal information and activate cognitive processes that are not particularly short (i.e. mere seconds), are necessary for thinking aloud. This was also confirmed by Veenman (2011; 2017), and Schellings (2011) added that automated tasks do not encourage thinking aloud, but complex ones do. Translating is an example of such a complex task, in which students have to simultaneously manage different processes (Göpferich, Jakobsen & Mees, 2009). Therefore, the Latin Buoyancy Task, in which students have to translate Latin sentences into Dutch, was a suitable task for thinking aloud.

Branch (2000) raised the concern that thinking aloud might be hindered when the cognitive load the task necessitates is too high. Cognitive load is the amount of simultaneous stimulation the working memory can successfully handle (Sweller, 1988). When a task is particularly complex, the cognitive load is likely to increase. Seeing the impossible nature of the second part of the Latin Buoyancy Task, and our aim to let participants perform it think aloud, this must be taken into consideration. The task and instrument, however, were designed to keep the cognitive load as low as possible – where possible. As the activity of translating Latin is not new to the participants, they are not complete novices at the task, which should mean that the cognitive load stays somewhat manageable (Schnotz & Kurschner, 2007). The cognitive load is further reduced by the use of complete wordlists for each part of the Latin Buoyancy Task. This meant that the participants did not have to use part of their working memory to recollect and remember word meanings. Moreover, the removal of time pressure is also known to decrease the cognitive load (de Jong, 2010). Participants in this study were allowed to spend as much time on the translations as they wanted, thus also limiting the pressure on their cognitive load. Furthermore, Wade (1990) concluded that participants with high cognitive functions are less likely to be over-asked whilst thinking aloud. As our participants belonged to the highest cognitive potentials in the Netherlands, overload seemed particularly unlikely. We, thus, expected that the cognitive load of our participants would not be over-asked nor hindering to the data collection.

Over the years, much attention has been paid to the validity of thinking aloud methods. Recently Zhang & Zhang (2019) addressed validity issues and as early as 1989 Russo, Johnson and Stephens raised the issue of possible reactivity. Reactivity refers to the idea that thinking aloud might affect the process, compared to what it would have been, if the participant had not been thinking aloud. Another form of reactivity relates to how the researcher might influence the participants' task process. However, a review study by Fox, Ericsson & Best (2011) that included 94 studies, found that reactivity is not an issue for *how* participants perform the task. However, participants do spend longer on performing the task when they think aloud, compared to when they perform the task silently (c.f. Brinkman, 1993; Ericsson & Simon, 1993; Leow & Morgan-Short, 2004; Bowles, 2010). To further ensure the validity of think aloud data, Ericsson and Simon (1993) advise researchers to follow three rules:

1. The researcher should remind participants to think aloud if they fall silent. To limit reactivity, this reminder should be short and not entail asking for clarification;
2. The researcher should have no other role than safeguarding the think aloud process, during the task performance;
3. The researcher should provide a warm-up exercise to ensure the participants are used to thinking-aloud.

Accordingly, validity issues with thinking aloud can and must be limited by the researcher as part of the design (c.f. Branch, 2000; Young, 2005).

Generally, thinking aloud can be implemented in different ways. Ericsson and Simon (1993) broke thinking aloud down into concurrent thinking aloud and retrospective thinking aloud. In the case of concurrent thinking aloud, participants give a live commentary to what they are doing. When the thinking aloud is retrospective, participants perform the task without thinking aloud, but explain their process after having completed the task. There are also examples to be found of prompted concurrent thinking aloud (e.g. Bannert & Mengelkamp, 2008). In these studies, the participants were asked to comment on what they were doing whilst performing the task, but only at specific intervals. We discarded this method of thinking aloud for this study as it might lead to prompting effects; by being asked what they are doing or having to choose from a list of strategies what they are doing, participants might be prompted to use a strategy they normally would not (Veenman, 2011). We did opt for a method in which continuous concurrent thinking aloud was combined with retrospective thinking aloud.

Concurrent thinking aloud had our preference, as it is not dependent on the participants' ability to recollect their process (Wade, 1990; Young, 2005). Another advantage of concurrent thinking aloud over retrospective is connected to cognitive load. As the load of a think aloud task should already be relatively high, it is unlikely that the participants can engage in socially desirable behaviour whilst performing the task (Young, 2005). Retrospective thinking aloud does provide participants with more room to display social desirability. However, despite its limitations, retrospective thinking aloud can provide extra insights into the participants' process. Unlike concurrent thinking aloud, it is not subject to task difficulty (Guan, Lee, Cuddihy & Ramey, 2006). By conducting the retrospective thinking aloud directly after performing the task, possible memory issues might be limited (Tomlinson, 1984; Kuusela & Paul, 2000). We also aimed at allowing the participants to elaborate about what they had done, why they had done that and how they were feeling throughout the task. Retrospective thinking aloud would enable us to do so. We believed this would support and strengthen our interpretation and coding of the data (c.f. Schellings, Van Hout-Wolters, Veenman & Meijer, 2013). Moreover, different methods are particularly helpful in gaining insights into learning strategies (Rogiers, Merchie & Van Keer, 2020) and increase the validity of thinking aloud (Young, 2005). Therefore, we made use of both concurrent and retrospective thinking aloud in which the data gathered by concurrent thinking aloud was supported by retrospective thinking aloud.

To further support the thinking aloud data, we included data collected from the questionnaires and the Emoji Chart (see Chapter 1). The data was analysed qualitatively and quantitively (see Chapters 4, 5 and 6). In other words, we implemented a mixed-methods design to come to a deeper understanding of academic buoyancy and strengthen the think aloud data (e.g. Tashakkori & Teddlie, 1998; Creswell, 2005; Morse & Niehaus, 2009).

An important aim of any think aloud research is to include the right participants, i.e. participants who will provide the richest information on the studied phenomena. Therefore, purposeful sampling is common practice in think aloud studies, as it enables researchers the possibility to select information-rich participants for their study (e.g. Patton, 1990; Patton, 2002; Creswell, 2009; Palinkas et al., 2015). By selecting participants that are of special interest for the research topic, more in-depth case analysis is possible, and despite relatively few participants, the data can be particularly enlightening (Patton, 1990; Creswell & Plano Clark, 2011). Patton (1990; 2002) described 15 different strategies for purposeful sampling and also added that these strategies can be combined. The best fitting purposeful sampling strategy depends on the study's aim.

The aim of our think aloud study was to establish differences between academic buoyant and non-buoyant students regarding how their cognitive, metacognitive and affective behaviour is influenced by a setback. To include both buoyant and non-buoyant students and thus study academic buoyancy as broadly as possibly, we opted for the sampling strategy labeled by Patton (1990; 2002) as deviant or extreme. Extreme sampling focusses on selecting and comparing participants that are on the extreme ends of the spectrum of the research topic (Patton, 1990; Patton, 2002; Palinkas et al., 2015). In our case, that would be those who we expected to be most and least academically buoyant. Or, in other words, the students who were most and least able to bounce back after a setback. By studying these students, we expected to be able to compare and contrast their behaviour and therefore form a clear picture of behaviour specific to non-buoyant students.

Extreme sampling also leads to a large variance within participants (Palinkas et al., 2015). A large variance within participants is crucial for the think aloud method, as it is labour-intensive and therefore the number of participants tends to be relatively small (Heirweg, De Smul, Devos & Van Keer, 2019). However, think aloud data collected is rich, so that small sample sizes suffice (Nielsen, 2000). This is reflected in previous think aloud studies involving a Latin translation task, with participants numbers ranging from two (Newland, 2016) to 30 (Eikeboom, 1970).[1] According to some, including five participants is sufficient in think aloud studies (Nielsen & Landauer, 1993; Nielsen, 2000). However, for the current study, we intended to include the relationship between independent and dependent variables and compare two heterogeneous groups of participants (i.e. buoyant and non-buoyant students). Therefore, both groups would have to include at least five participants (Macefield, 2009). Thus, we would include at least ten participants, of which at least five were expected to be particularly buoyant and at least five to be particularly non-buoyant.

Not only the number of participants is relevant in the case of extreme sampling, but also factoring in which characteristics are important to include within the sample (Patton, 2002). For answering our research questions, the most important characteristic of our participants was that they were either particularly buoyant or non-buoyant. To determine this, we made use of the Anagram Task (see Chapter 1, p. 41). How we made use of this task to ascertain the extent of buoyancy in students is explained in full in section 2.2. Besides buoyancy, we included two other characteristics in the sample; [1] years of learning Latin and

---

[1] *For other participant numbers included in think aloud studies regarding Latin translation the following applies: Van Krieken, 1982: 10 (concurrent); Van Houdt, 2008: 28; Florian, 2015: 12 (concurrent); Karten, 2015: 11 (concurrent); Boyd, 2018: 12 (retrospective); Luger, 2020: 18 (retrospective).*

[2] giftedness. In the following paragraphs we first divulge how we come to choosing third year Latin students as a student characteristic and then we explain how we operationalised the complex concept of giftedness.

We had three reasons to focus on students in their third year of Latin. The first was that previous studies related to Latin translation, particularly in the Netherlands, have focused on students in the upper years of high school (e.g. Eikeboom, 1970, van Krieken, 1982; Sarkissian, 2008; Florian, 2015; Luger, 2020). Therefore, there seemed an opportunity to fill in a gap by studying younger students. Secondly, third year students have usually already been introduced to some of the more complex morphology and syntax of Latin, providing us with the possibility to create sentences for the pre- and post-setback tasks that cannot be translated at first sight. This was necessary as not to impede the thinking aloud (c.f. Payne, 1994; Schellings, 2011; Veenman 2011; Veenman, 2017). Moreover, the larger complexity of the translatable sentences would ensure the untranslatable task is not too obviously untranslatable (see also design principles 4 and 5, Chapter 1). The third and final reason to focus on third year Latin students was their age. For, generally, when students are older than 12, their metacognitive skills become more transferable between different domains and less domain specific. As the effect of the setback on metacognitive strategies is one of our focusses, this seems a particular fitting age for our participants. Thus, being third year Latin students became one of the characteristics for our participants.

Giftedness was the final of our characteristics. This characteristic came from our particular interest in giftedness and buoyancy. Gifted students are not used to being cognitively challenged and, therefore, have relatively little experience with challenges. This lack of experience in overcoming challenges can lead to multiple problems and might for a part explain why so many gifted students end up underachieving (Balduf, 2009; for more, see Introduction). This led to giftedness being included as a participant characteristic.

By including giftedness as a characteristic, we posed ourselves with a conundrum. Despite no lack of attempts, there is still no broadly accepted definition of giftedness (e.g. Renzulli, 1986; Betts & Neihart, 1988; Ziegler & Heller, 2000; Gagné, 2003; Cramond, 2004; Renzulli, 2016; Papadopoulos, 2020). Nevertheless, we still needed a method of operationalising 'giftedness' for our study if we were to include it as a characteristic for screening. Making use of an IQ-test score of TIQ = >129 might seem a logical method, particularly as many studies related to giftedness do make use of IQ-scores to define giftedness (e.g. Bennett-Rappell & Northcote, 2016; Makel, Kell, Lubinski, Putallaz & Benbow, 2016). Besides the ongoing debate that IQ-score alone does not define giftedness (Papadopoulos, 2020), using an IQ-test was in our eyes not a valid option for multiple reasons.

The first reason was that at the time of our data-collection (spring 2017), Dutch psychologists were still having to make use of the outdated WISC-III as the mainstream IQ-test. This was due to a lack of a Dutch version of the WISC-IV and the WISC-V translation was still pending. Some practitioners even resorted to using the RAKIT-2 to establish giftedness, an IQ-test that was developed with particularly children with low verbal capacities in mind (Dek & Kooij, 2012). Depending on the used test, the TIQ could differ substantially (Ruiter, Hurks, Timmerman, 2017). More importantly, an IQ-test is only a momentary snapshot of a child's capabilities (Christie, 2005). Finally, not all students take an IQ-test, and, in the Netherlands at least, an IQ-test is often only advised when there is a (behavioural) problem in the classroom. This in turn means that only certain gifted students have a known IQ-score, in particular those who Betts & Neihart (2010) either classed as 'challenging' or 'twice exceptional'. Given all these reasons, an IQ-score, could not be a leading prerequisite for our study.

Being a Latin student *was* a prerequisite for our study, as our task involved translating Latin (for more on the domain choice, see Introduction, p. 6). This presented us with a possibility to operationalise 'giftedness', without an IQ-test. In the Netherlands, secondary education is split into distinct levels: pre-vocational education, higher general education and pre-university education. Latin is only offered as part of the pre-university curriculum and even then, only as a non-compulsory subject. Schools offer Latin to students who can cognitively handle 'something more' (c.f. van der Plaat, 2016). Latin is, thus, in general, only offered to high potentials. Therefore, by selecting students with Latin in their curriculum we could be certain that they had at least more than average cognitive potential. This rationale was added to by only inviting Latin students who attended a *Begaafdheidsprofielschool* as our think aloud participants. These schools tend to have large bodies of so-called gifted students, due to their special gifted programs. By following this method instead of a strict (non-existent) definition of giftedness, we expected to gain insights into academic buoyancy in a group of gifted students.

In sum, this Chapter focusses on how the data for this study relating to academic buoyancy was collected. We applied a mixed-methods approach, in which thinking aloud was central. Data collected via continuous concurrent thinking aloud was supported with data collected from retrospective thinking aloud and questionnaires. To ensure our participants were gifted and included both academic buoyant and non-buoyant students, we made use of extreme sampling amongst the highest academic potentials in the Netherlands. How these methods were implemented is the focus of the following sections.

## 2 METHOD

### *2.1 Participants*

From a group of 83 third year Latin students, we selected students for a think aloud study using extreme sampling. We originally approached 18 to participate in the study. However, for one student we did not receive consent to participate in the study and due to audio problems, we were unable to transcribe the recordings of another student. The study thus eventually included 16 participants. The students attended the same four schools as the students who participated in the testing of the different instruments (see Chapter 1, p. 33). The selected students were aged between 14 and 15 and all attended a *Begaafdheidsprofielschool*. Despite an IQ test not being a stipulation for participating, 11 of the students were known by their teacher to have previously taken an IQ test, in all cases resulting in a TIQ >129. Ethics approval was obtained from the university board prior to participant selection and parents provided active consent. Students received no incentive for participation.

#### 2.1.1 Participant screening

As presented in Chapter 1, we designed the Anagram Task as a task-based instrument to determine the extent of academic buoyancy in students. Based on the testing of the Anagram Task and it's overlap with the Latin Buoyancy Task, we determined that we would need about 85 students to perform the Anagram Task, if we wanted a mix of students who were particularly buoyant or non-buoyant. We came to this as when comparing the results of 85 students from the Anagram Task to those from the Latin Buoyancy Task, we found 11 students were unable to bounce back after the setback on both tasks and only two mishits (see Chapter 1, p. 46). We, therefore, after obtaining passive consent from all the parents, distributed the Anagram Task Version A amongst four classes and a total of 83 students performed the Anagram Task.

The data was collected at the students' schools by their form teacher, six to eight weeks prior to the first think aloud data being collected. No significant incidents were reported to have occurred during the data-collection. The teacher explained that the Anagram task was part of a study related to linguistic problem-solving and then demonstrated what anagrams are and how to solve them. Following this, students were given several minutes to practice solving four lettered anagrams. Finally, the students received the three subtasks of the Anagram task successively. Before starting each subtask, it was emphasized that the anagrams did not have to be solved in order. The students received two minutes to solve as many anagrams as they could for each subtask. The debriefing session was different to the testing phase of the instrument, as the students were not told about the setback task and its purpose. The students could not be informed about this, as such knowledge was likely to impact the effect of the Latin Buoyancy Task as the tasks' designs were similar in nature.

To determine which students were the most and least non-buoyant, we compared their pre-setback scores to their post-setback ones. The internal consistency of the pre-setback task was .70. (compared to .68 in the testing phase, see Chapter 1, p. 55). Pre-setback, the students found an average of 6.02 solutions ($sd$ = 1.80). For the third set they found 5.07 ($sd$ = 1.83) solutions. The students thus found .95 ($sd$ =2.07) fewer solutions in the third task compared to the first ($p$ < .001). The scores of 20 students decreased by more than a standard deviation. These students were interpreted as being non-buoyant on the anagram task and, thus, also likely to be non-buoyant on the Latin Buoyancy Task as well. On the other end of the spectrum, seven students not only bounced back to their pre-setback score, but improved it. These students were deemed to also be buoyant on the Latin Buoyancy Task.

As we had decided that extreme sampling would provide us with the richest data, we invited the eleven students who had the largest decline post-setback (i.e. non-buoyant students on the anagram task) to participate in the study. For all these students we received active consent for participation from the parents. However, due to audio-problems it was not possible to fully transcribe the recordings of one student into a verbal protocol. The data from this student was therefore not included in the study, leaving us with ten non-buoyant participants Besides these participants, we also invited all seven students who improved on their scores post-setback (i.e. buoyant students on the anagram task). One of these students was the student whose parents did not consent to participation in the study. Therefore, we were left with six buoyant participants. These participants and their synonyms are presented in Table 2.1.

Finally, we also invited one student to participate in a trial run of the thinking aloud. This participant was selected at random from the participants who had performed the anagram task, but had not been selected to participate in the study itself as her results did not fall into the extremities of significant improvement or significant deterioration. For the purposes of this dissertation, we shall call this student Victoria.

*Table 2.1 Buoyant and non-buoyant participants*

| Buoyant | Non-buoyant |
|---|---|
| Flavia | Anthony |
| Horatia | Bella |
| Livia | Claudia |
| Marcus | Diana |
| Nona | Ennius |
| Phaedra | Gaius |
| | Julia |
| | Octavia |
| | Quintina |
| | Rufus |

### *2.2 Materials*

The central task for the thinking aloud was the Latin Buoyancy Task (Appendix 1.1). The version we made use of included the final adjustments of the *equus* sentence as discussed in Chapter 1 (p. 38). Besides the thinking aloud data, we also collected data using the Emoji Chart and the Mindset and Frustration Tolerance Questionnaires, (see Chapter 1). The final material we made use of was a warm-up task for the thinking aloud. The warm-up task aimed at stimulating the students' ability to perform a task whilst thinking aloud (Ericsson & Simon, 1993; Young 2005). For the task itself, a moderate difficulty sudoku puzzle was chosen. This task was chosen as it was a non-linguistic task that involved problem-solving and was familiar to students. The sudoku was retrieved from internet and can be found in Appendix 2.1.

### *2.3 Procedures*

In this section we present all the procedures we followed during the data collection. First, we present the procedures regarding the Mindset and Frustration Tolerance Questionnaires, then we present the trial run of the thinking aloud. This is followed by the procedures of the concurrent thinking aloud, the retrospective thinking aloud and the debrief. In total the whole process of the introductions, warm-up task, concurrent thinking aloud, retrospective thinking aloud and debriefing took on average about an hour for each participant. The longest time needed for one participant to go through the whole procedure was just under 90 minutes. For the procedures regarding the creation of the verbal transcripts and their coding, we refer to Chapter 3.

### 2.3.1 Questionnaires

All the students who performed the Anagram task, also filled in the Mindset and Frustration Tolerance Questionnaires. They answered the questionnaires after having completed the Anagram Task. Consequently, the questionnaires were conducted in a classroom by the students' form teacher. The mindset and frustration tolerance items were combined to form one questionnaire, which was distributed on paper. The teacher then asked the students to highlight per statement to what extent they agreed with it. The students were given 15 minutes to fill out the questionnaire. At the end, the students were asked to check if they had answered all the questions and if they had done so using 1 to mark total agreement and 6 to mark total disagreement.

### 2.3.2 Trial run of the thinking aloud

Before carrying out the think aloud process, a trial run was conducted, as we felt it was essential for the researcher to practice the think aloud process with a student, Victoria. The aim of this trial run was to test the materials one more time and for the researcher to gain experience in attending a think aloud process and, therefore, to make interference from the researcher's presence less likely on the participants' translation process.

The trial run was set up similarly to the actual think aloud process: it was at Victoria's school in a one-on-one setting with the researcher during a free period and it was recorded for evaluation of the researcher. The camera was set up directly in front of the student and mainly focused on their work. The researcher sat next to Victoria. We told Victoria that she was participating in a trial run of the Latin Buoyancy Task and asked her to complete the translation task as best she could in as much time as she needed. We also informed Victoria that the researcher was only present to help the her remember to think aloud and could not answer questions regarding the content of the task (cf. Ericsson & Simon, 1993; Young, 2005). When Victoria was ready, she started the thinking aloud warm-up task (idem). After about ten minutes, Victoria had completed the sudoku. The thinking aloud was going well, so with no further instructions on thinking aloud, we moved on to the Latin Buoyancy Task.

To start, Victoria was asked to put a '1' in the box of the emoji from the Emoji chart that best described her mood at that moment. Then we proceeded to the pre-setback translation task. Each subtask had been printed alongside the accompanying word list separately, with space for Victoria to write down her translation. During the task, the researcher's only role was to prompt thinking aloud when the student fell silent for longer than two seconds. To avoid prompting effects these reminders were short, the most used being 'Aloud'[1] (Ericsson & Simon, 1993; Veenman, 2011). Whilst Victoria was thinking aloud, the researcher made mental notes of any pivotal moments, so to be able to ask about those moments during the retrospective thinking aloud. This was not done on paper, as that might be visible to the student and inadvertently influence their process. After having completed each of the three sentences Victoria filled in the Emoji Chart again, the second time with a '2' and so on. Victoria also gave a grade out of ten representing how correct she felt her translation of each subtask. So, by the end of the task we had collected three sets of translations, three subjective marks and four emoji.

The researcher then preceded to the retrospective thinking aloud. Victoria was shown each subset individually and asked how she had arrived at the translation. Victoria for example told us that combining the word meanings to make a logical sounding sentence was her main strategy. We also asked how she had come to the subjective grades for each translation and which of the subtasks she felt was the easiest and which the most difficult. Victoria felt the second task was the most difficult, but felt that because she had done her best she should have at least a decent mark for the other two subsets.

Finally, we debriefed Victoria about the actual topic of the study and the untranslatable nature of the second translation task. Victoria laughed and said she recognised that when a task is too difficult, she sometimes gives up easier. She was interested in what the researcher thought of her translation process. After having asked her for feedback on the role of the researcher, ("place the camera more at a sideward angle not directly in front of the participants, to be less distracting and picking just one emoji is quite difficult") we thanked Victoria thanked and asked her not to share any particulars from the study with any classmates, in case any of them were participating in the study. Victoria promised not to do so and then the trial run was ended.

The recordings of the trial run were shared with members from the researchers' research group. They provided feedback regarding the prompting. Their advice was to prepare a more varied selection of short but friendly prompts besides 'aloud' for when the participants fell silent. Thus, for example 'What are you thinking now?'[2] and 'Share your thoughts'[3] were added to the repertoire of

[1] *Hardop.*
[2] *Wat denk je nu?*
[3] *Deel je gedachtes maar.*

the researcher for the think aloud process proper. With this feedback gained, we were confidently prepared to start the proper data collection.

### 2.3.3 Concurrent think aloud processes

For the think aloud processes proper, many of the same procedures were followed as during the trial run. The researcher collected the data in a one-on-one situation with the participants at their schools over a period of three weeks. The sessions were conducted within the regular school day hours, either during a Latin lesson or during a free period. All sessions were recorded on video, with the camera being positioned to record the participants' progressing translation. The camera was not placed directly in front of the participants, but at an angle, based on the feedback from the trial run student. Again, the researcher sat next to the participants and only made mental notes.

After introductions, the participants were told that the researcher was doing a study into students' translation processes to be able to improve Latin translation instruction. The researcher then gave the students the following instruction regarding the think aloud task:

> "In a moment you are going to translate three sets of three sentences from Latin to Dutch, so a total of nine sentences. It is important that while you are doing so, you say aloud what you are thinking. The aim is that you share all your thoughts, so also those that you do not think are relevant or those that are not about the task. To help you with the thinking aloud, we will do a practice task first. If you forget to think aloud, I will remind you to do so. An alphabetized wordlist will be supplied with each set to help you. My role today is merely to help remind you to think aloud and make sure the tasks are completed in the right order. I can therefore only answer questions about the procedure and no other questions for example about the Latin. I would also like to ask you to translate the sentences as well as you can. It is not an exam and I will not share any observations with your teacher, but you will help the research if you translate as well as you are able to. Finally, it is good to know that there is no time limit, so you can take as long as you like. When you have completed the translations, we shall have a short chat about how you approached the tasks." [1]

[1] *Zo meteen ga je drie setjes van drie zinnen van het Latijn naar het Nederlands vertalen, in totaal dus negen zinnen. Het is belangrijk dat terwijl je dit doet, je hardop zegt wat je allemaal denkt. Het is de bedoeling dat je alle gedachtes die je hebt hardop deelt, dus ook de gedachtes die jij niet relevant vindt of die niet gaan over de taak. Om je te helpen met het hardop denken gaan we eerst een oefentaak maken en als je vergeet hardop te denken zal ik je eraan herinneren het alsnog te doen. Bij elk setje is een alfabetische woordenlijst toegevoegd om je verder te helpen. Mijn rol vandaag is om jou alleen te helpen herinneren aan het hardop denken en het volbrengen van de taken op de juiste volgorde. Ik kan dus alleen vragen over de procedure beantwoorden en je inhoudelijk niet helpen als je vragen hebt over bijvoorbeeld het Latijn. Ik wil je verder vragen zo goed mogelijk de zinnen te proberen te vertalen, het is geen toets en ik zal geen observaties met*

The researcher then asked whether the participants had any questions. If they did, these were answered first, before beginning the warm-up task.

Most participants completed the warm-up task within about ten minutes. We then moved on to the filling in the Emoji Chart for the first time. Based on the feedback from the trial run, we told the participants to pick the emoji which was most fitting and that if they could not choose just one, to highlight two or a maximum of three. The participants then set about the pre-setback translation task, whilst thinking aloud. Most participants made notes on the paper with the sentences, and all wrote their translations down in the designated area for the translations at the bottom of the page. When they indicated that they were done, the researcher asked them to 'Write down on the top of the page, which mark you would have expected for these translations if it had been an exam'.[1] Before moving on to the next sentences, the participant filled in the Emoji Chart again and this whole process was then repeated for the setback task and the post-setback task.

During the Latin Buoyancy Task, the necessary prompting by the researcher varied. Most participants did not often fall silent for longer than a few seconds, and thus needed little prompting, particularly towards the end of the task. One of the participants (Diana), however, struggled throughout the task to think aloud and required frequent prompting. Even with the prompting it was sometimes arduous, with the participant for example saying that her head was empty. Apart from this, there were no other noteworthy incidents during the data collection.

### 2.3.4 Retrospective think aloud processes

Just as during the trial run, the retrospective think aloud process followed directly after the participant had highlighted an emoji for the fourth time. The participant's written work (their notes and translations) served as the prompt for the retrospective thinking aloud. The researcher and participant went through the work, task by task. This was done semi-structured: each participant was asked how they had approached the task, what difficulties they had encountered, how they had solved those difficulties, if they had experienced any frustrations and how they came to their expected marks. Depending on the participants' answers,

---

*je docent delen, maar je helpt het onderzoek wel als je het zo goed als je kunt probeert te vertalen. Tot slot is het goed om te weten dat er geen tijdslimiet is, dus je mag er zo lang over doen als je wil. Na afloop van het vertalen, zullen we nog even kort nabespreken hoe je de taken hebt aangepakt."*

[1] *Schrijf boven je vertaling het cijfer dat je zou verwachten voor deze vertalingen als dit een toets zou zijn geweest.*

the researcher might have asked clarifying questions, particularly when the participant used subjective words (c.f. Lang & Molen, 1998).

After the Latin Buoyancy Task had been fully talked through, the researcher showed the participants their filled-out Mindset Questionnaire and asked why the students had filled in the questionnaire as they had. The aim of this was to further validate the Dutch Mindset Questionnaire and help explain certain deviating outcomes compared to Dweck's original (see Chapter 1, p. 60).

### 2.3.5 Debrief

When the retrospective thinking aloud was complete, the researcher began the debrief. The aim of this debrief was that no participant left feeling particularly negative about their performance due to the setback task. During the debrief all but one of the participants were told the following:

> "I want to thank you for participating today, you have provided a lot of information. I do, however, have a little confession to make. The second set of sentences you had to translate, were actually nonsense Latin and could therefore not be translated. [Pause for participants' first reaction] I am specifically interested in what happens to students' translation processes when they get stuck. There is a theory that some students totally give up when they come across a problem, whilst others are spurred on by problems. So, I intentionally gave you an untranslatable task, to see if that theory holds true for students when they are translating. I want to thank you again for participating." [1]

All participants were quite amused by the fact that the setback task was untranslatable. Most participants confessed to being relieved and exclaimed things like 'it was not down to me'[2]. Nearly all the participants were intrigued by the buoyancy theory and asked questions about it. In some cases, they started talking about how they recognised themselves in the theory or asked the researcher what she had seen in their process. Interestingly, four of the participants asked for specific tips to improve their translation skills. Each participant was asked not to share any particulars about the task with other students in their class. All agreed to this and often noted that it was necessary

---

[1] *Ik wil je in ieder geval erg bedanken voor het meedoen vandaag, je hebt veel laten zien. Ik moet nog wel iets bekennen. De tweede set van drie zinnen die je moest vertalen, waren eigenlijk onzin Latijn en konden dus helemaal niet vertaald worden. [Pause for participants' first reaction] Ik ben namelijk specifiek geïnteresseerd in wat er gebeurt met de vertaalstrategie van leerlingen op het moment dat ze vastlopen. Er is namelijk een theorie waaruit zou blijken dat sommige leerlingen als een ding niet helemaal goed lukt, de rest een beetje opgeven en anderen juist aangemoedigd worden het extra goed te gaan doen daardoor. Dus, ik heb expres een onmogelijke vertaaltaak voorgelegd, om te kijken of die theorie geldt voor leerlingen die aan het vertalen zijn. Ik wil je in ieder geval echt heel erg bedanken.*

[2] *Het lag niet aan mij.*

for the study's design that the participants were unaware of the full story, or that they hoped their friend would be completely taken in by it. When all the students' questions had been answered, the session was ended.

One participant (Livia) received an alternative debrief. Livia was the only participant who gave herself a very high mark for the setback task and said that she was particularly content with how the setback task went. The researcher did not want to unsettle her by saying that task was untranslatable. Therefore, the debriefing this student received was slightly adapted. Instead of specifically saying that the second task was fully untranslatable, the researcher told Livia that it 'contained untranslatable parts'. (For more on this particular participant, see Chapter 5 and 6) The rest of the debrief was the same as that of the other participants.

### 2.3.6 Other processes related to the data collection

The participants' expected academic buoyancy had been determined by how they performed on the Anagram Task. Besides the think aloud data (including their Emoji Chart and subjective scores) and data from the questionnaires, we derived five other types of data from the think aloud process:

1. Buoyancy in accuracy: i.e. how accurately the participants had translated the post-setback task compared to the pre-setback task;
2. Translation proficiency: i.e. how proficient the participants were at translating Latin into Dutch during the pre-setback task;
3. Time on task: i.e. how long the participants spent on the Latin Buoyancy Task;
4. Number of switches: i.e. how many times the participants switched between different thinking-activities during the Latin Buoyancy Task;
5. Average thinking activity duration: i.e. the average duration of each individual thinking-activity.

## *2.4 Analyses*

The participants were all given a random number from one to sixteen, to be able to conduct the analysis anonymously. To bring the participants behind the data more alive in the descriptions, we gave the participants alphabetical Latin pseudonyms. The Mindset and Frustration Tolerance Questionnaires and the verbal transcripts were analysed using different methods to provide data that could be further analysed qualitatively and quantitively. These methods and what data they yielded are described in the following sections.

### 2.4.1 Questionnaires

The internal consistency was calculated for the Mindset and Frustration Tolerance Questionnaire items separately. To do so, the answers of all 83 students were included, not just the answers of the 16 participants.

For the Mindset Questionnaire the analysis was the same as presented in Chapter 1: For each participant, the mean score was calculated from the three items. When the mean was lower or equal to three, this implied the student was more inclined towards a fixed mindset preference. If the mean score was four or higher, the student was classed as being more inclined towards a growth mindset preference. Any score that fell between these boundaries was taken as an indication of the student not having a clear mindset preference (e.g. Hong, et al., 1999).

For the frustration tolerance items, the mean was also calculated. As described in Chapter 1, when the mean was higher than 4.0, the participant was interpreted as having a relatively high frustration tolerance. When the mean was lower than 3.0, this was interpreted as the participant having a relatively low frustration tolerance. Any mean that did not fall in either of the two categories was understood as the participant neither leaning towards a particularly low or high frustration tolerance (Wright, Lam and Brown, 2009).

### 2.4.2 Accuracy and Translation proficiency

The participants had written down their translations of all the sentences. In the case of the pre- and post-setback task, these translations were scored by two raters. In a single case, the participant had said something different than they had written down as their translation. In that case, the participants' written work was leading. For the rating we made use of the same scale as in the design cycles: two points for a perfectly translated sentence, one for a translation containing a minor mistake, and nil for a translation with more than one mistake. Thus, the pre- and post-setback tasks were worth a maximum of six points each. For an example of this scoring system, see Chapter 1, p. 25. Both raters were accustomed to rating Latin translations and they agreed about the scoring of all the translations.

We determined the participants' proficiency in Latin translation tasks through the accuracy scores of the pre-setback task. The participants had translated those sentences under normal circumstances, thus giving a picture of their ability to translate Latin to Dutch. We did not make use of the participants' school grades to determine their translation proficiency as these grades also represent other skills. Making use of a school grade for a translation test also did not suffice. For, the translation tests would not have been the same for all participants, and we wanted a way of testing their proficiency that was equal for all the participants. Therefore, we used their scores from the pre-setback task as an indication of their translation proficiency. When the participant had produced a perfect translation, this participant was classed as high proficient, whereas participants who scored two points or less were classed as low proficient.

### 2.4.3 Time on task

The recordings made it possible to collect time codes relating to how long the participants spent on the Latin Buoyancy Task. By playing back the recordings and stopping them at the moment the participant says they had completed any of the tasks, we were able to note in seconds how long the participants had spent on not only the Latin Buoyancy Task as a whole, but also on the pre-setback task, the setback task and the post-setback task individually.

### 2.4.4 Number of switches

The segmentation of the verbal transcripts (see Chapter 3) enabled us to measure the number of times the participants switched from one think-activity to another. The number of switches was calculated by subtracting one from the total of thinking activities displayed by that participant. We calculated the number of switches separately for the pre-setback, the setback and the post-setback task.

### 2.4.5 Average thinking activity duration

We also determined the average duration of thinking activities. Again, we did this for the pre-setback, the setback and the post-setback task separately. To do so, the total amount of time in seconds which the participant spent on each subtask, was divided by the total number of thinking-activities used in that subtask.

The above is summarized in Table 2.2. This table for example demonstrates that the Anagram Task was used to select participants. On the other hand, the verbal transcripts of the pre-setback task were used to gather data relating to the participants' proficiency, how long they spent on translating it, how often they switched from one thinking activity to another and the average duration of said thinking activities.

## 3 FIRST OVERVIEW OF RESULTS

As explained in the section on the participant screening, 16 selected participants were included in the think aloud study. All the data gathered relating to these participants can be found in Table 2.3. In the coming sections we present these results in three parts. First, information is given regarding the participants themselves. Then, the results relating to the questionnaire are elucidated. Finally, the results relating to the Latin Buoyancy Task are presented.

*Table 2.2 Summary of which materials were consulted to gain that type of data*

| | | Expected academic buoyancy | Academic buoyancy Latin | Proficiency | Mindset | Frustration tolerance | Time on task | # of switches | T-A[1] duration |
|---|---|---|---|---|---|---|---|---|---|
| Anagram Task | | x | | | | | | | |
| Questionnaire | | | | | x | x | | | |
| Latin Buoyancy Task | Translations & Transcripts pre-setback | | x | x | | | x | x | x |
| | Transcripts Setback | | | | | | x | x | x |
| | Translations & Transcripts post-setback | | | | | | x | x | x |

*Note.* [1]*T-A stands for thinking-activity.*

*Table 2.3 Results per participant*

| | Participant info | | | Questionnaires | | Latin Buoyancy Task | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Student | Pseudonym | Anagram[1] | IQ test[2] | *M* Mindset | *M* Frustration | Translation Accuracy | | | Time on Task[3] | | | Number of Switches | | | Average TA duration[4] | | | |
| | | | | | | Pre[5] | Post[6] | DS[7] | Pre | SB[8] | Post | Pre | SB | Post | Pre | SB | Post | |
| 1 | Anthony | -5 | Yes | 2.67 | 3.38 | 6 | 1 | -5 | 779 | 776 | 450 | 49 | 41 | 17 | 15.9 | 18.93 | 26.47 | |
| 2 | Bella | -3 | Yes | 4.33 | 3.38 | 3 | 1 | -2 | 397 | 644 | 448 | 33 | 37 | 24 | 12.03 | 17.41 | 18.67 | |
| 3 | Claudia | -4 | Yes | 4 | 3.88 | 4 | 3 | -1 | 400 | 504 | 283 | 32 | 39 | 24 | 12.50 | 12.92 | 11.79 | |
| 4 | Diana | -3 | No | 4.67 | 5.50 | 4 | 1 | -3 | 823 | 532 | 534 | 51 | 24 | 29 | 16.14 | 22.17 | 18.41 | |
| 5 | Ennius | -5 | No | 5 | 3.63 | 6 | 1 | -5 | 484 | 480 | 291 | 30 | 32 | 12 | 16.13 | 15 | 24.25 | |
| 6 | Flavia | +2 | Yes | 2.33 | 4 | 2 | 1 | -1 | 570 | 779 | 287 | 48 | 53 | 12 | 11.88 | 14.70 | 23.92 | |
| 7 | Gaius | -3 | Yes | 3 | 3.63 | 6 | 3 | -3 | 361 | 782 | 260 | 33 | 55 | 17 | 10.94 | 14.22 | 15.29 | |
| 8 | Horatia | +2 | Yes | 5.33 | 1.63 | 5 | 3 | -2 | 517 | 734 | 223 | 23 | 31 | 13 | 22.48 | 23.68 | 17.15 | |
| 9 | Julia | -6 | No | 5.33 | 3 | 3 | 0 | -3 | 631 | 480 | 450 | 42 | 27 | 21 | 15.02 | 17.78 | 21.43 | |
| 10 | Livia | +3 | Yes | 5 | 3.38 | 4 | 3 | -1 | 1013 | 689 | 440 | 35 | 23 | 19 | 28.94 | 29.96 | 23.16 | |
| 11 | Marcus | +4 | Yes | 3.33 | 3.64 | 2 | 3 | +1 | 669 | 931 | 211 | 66 | 67 | 18 | 10.14 | 13.90 | 11.72 | |
| 12 | Nona | +2 | Yes | 2.33 | 4.25 | 4 | 2 | -2 | 453 | 572 | 305 | 22 | 44 | 22 | 20.59 | 13 | 13.86 | |
| 13 | Octavia | -5 | Yes | 4.33 | 4.50 | 4 | 0 | -4 | 342 | 539 | 225 | 19 | 51 | 20 | 18 | 10.57 | 11.25 | |
| 14 | Phaedra | +6 | Yes | 5.33 | 1.63 | 3 | 1 | -2 | 505 | 758 | 582 | 25 | 33 | 28 | 20.20 | 22.97 | 20.79 | |
| 15 | Quintina | -4 | No | 3.70 | 3.60 | 1 | 3 | +2 | 345 | 562 | 499 | 15 | 18 | 20 | 23 | 31.22 | 24.95 | |
| 16 | Rufus | -4 | No | 3 | 4.50 | 4 | 0 | -4 | 953 | 660 | 492 | 59 | 59 | 47 | 16.15 | 11.59 | 10.47 | |
| *Mean* | | -1.44 | | 4.01 | 3.59 | 4 | 1.64 | -2.19 | 552.6 | 650.8 | 365.87 | 36.38 | 39.63 | 21.44 | 16.88 | 18.13 | 18.35 | |
| *SD* | | 3.88 | | 1.20 | 0.97 | 1.36 | 1.15 | 1.94 | 197.51 | 138.45 | 124.48 | 14.8 | 14.19 | 8.48 | 5.13 | 6.28 | 5.44 | |

*Notes.* [1] *This column includes how the participants performed the post-setback Anagram Task compared to the pre-setback task;* [2] *This column includes whether the participant was known to have taken an IQ-test according to their teacher, all participants who had (indicated by 'yes'), had been found to have a TIQ >130;* [3] *This column includes the time the participants spent on the (sub)task in seconds;* [4] *This column includes the average duration of the thinking activities in seconds;* [5] *Pre stands for pre-setback task;* [6] *Post stands for post-setback task;* [7] *DS stands for difference in score, i.e. the difference in accuracy between the pre- and post-setback task;* [8] *SB stands for setback task.*

### 3.1 Participant information

The pseudonyms of the 16 participants are presented in the second column of Table 2.3. These pseudonyms also demonstrate the gender of the students: any name ending in -a denotes a female participant, whereas all other ending denote a male participant.

The six students who were invited as participants for the study due to their display of academic buoyancy during the Anagram Task, were the students who in Table 2.3 had positive scores in the third column. Thus, Flavia, Horatia, Livia, Marcus, Nona and Phaedra were expected to be academic buoyant participants based on their results from the Anagram Task. These participants had all improved upon their pre-setback anagram score after the setback. All other participants' scores had deteriorated the most of the 83 students. These ten students thus fulfilled the role of the non-buoyant participants in this study.

As becomes clear from the fourth column, only Diana, Ennius, Julia, Quintina and Rufus had not partaken in an IQ-test, according to the knowledge of their form teacher. All other participants had done so and were found to have a TIQ of >129.

### 3.2 Questionnaires

The internal consistency of the Mindset Questionnaire items was .74. For further information on the scores relating to all 83 students who filled out the Mindset Questionnaire (see Chapter 1, p. 56). The average mindset score reported by the 16 participants was 4.01 (*sd* = 1.2). The results from five participants (Anthony, Flavia, Gaius, Nona and Rufus) indicated that they had a fixed mindset preference. Marcus and Quintina reported no particular mindset preference. The scores of the remaining nine participants indicated a growth mindset preference.

In the case of the Frustration Tolerance items, the internal consistency was .83. Again, for further information on the scores relating to all 83 students (see, Chapter 1, p. 64). On average, the 16 think aloud participants scored 3.59 (*sd* = 0.97) on the Frustration tolerance items. Diana in particularly reported having a high frustration tolerance, but also Flavia, Nona, Octavia and Rufus reported being able to deal with frustration well. On the other end of the spectrum Horatia and Phaedra indicated having a low frustration tolerance.

### 3.3 Latin Buoyancy Task

On average, the participants scored four out of the available six points for their translations of the pre-setback task (*sd* = 1.36). Anthony, Ennius, and Gaius even achieved perfect translations. These participants were thus classed as high proficient translators, alongside Horatia, who scored five points. On the other end of the scale, Flavia, Marcus and Quintina only scored a maximum of two points, indicating that in this group they were low proficient translators.

In the case of the post-setback task, we saw that the average accuracy decreased from 4.0 (*sd* = 1.36) to 1.64 (*sd* = 1.15). Julia, Octavia and Rufus all scored nothing in the post-setback task. High proficient Anthony and Ennius both only managed to score one point this time around for their translation. On the other hand, Marcus and Quintinia improved upon their pre-setback score.

The participants spent an average of just under 26 and a half minutes on translating the whole Latin Buoyancy Task (*sd* = 5.19 minutes, range = 18.26-35.04 minutes). On average, they spent longest (650.8 seconds, *sd* = 197,51) on the setback task, and the shortest amount of time on the post-setback task (365.87 seconds, *sd* =138.45). Three participants spent particularly long on the pre-setback task. For, Diana, Livia and Rufus all spent more than 13 minutes on the task.

Regarding switches made between thinking activities, we found that, on average, the number of switches slightly increased from the pre-setback task (*M* = 36.38; *sd* = 14.8) into the setback task (*M* = 39.63; *sd* =14.19). This was followed by seemingly large decrease in the post-setback task (*M* = 21.44; *sd* = 8.48).

There was a much smaller change visible on average regarding the average length of each thinking activity throughout the Latin Buoyancy Task (pre-setback: *M* = 16.88; *sd* = 5.13 / setback: *M* = 18.13; *sd* = 6.28 / post-setback: *M* = 18.35; *sd* = 5.44).

## 4 DISCUSSION

The results presented above are quantitively analysed and discussed in Chapter 4. Chapters 5 and 6 provide a qualitative approach and sequential discussion of the findings. The current Discussion section is limited to the actual method of data-collection. The consequences of our methodological choices, such as extreme sampling and our operationalisation of giftedness, are considered here. Also, the thinking aloud process itself is reflected on. Finally, the relationship between the different methods used to gather data in this mixed-methods design is discussed. We look at whether all the different methods were necessary and how the methods supported or hindered each other.

Due to the extreme sampling, we only included the students who demonstrated either the most or least extent of academic buoyancy during the Anagram Task as participants. We were somewhat limited in the number of participants, particularly with regards to buoyant participants. This was aggravated by the loss of an intended buoyant participant due to a lack of permission for participation. However, having six buoyant participants was still above the threshold of five (Nielsen & Landauer, 1993; Nielsen, 2000; Macefield, 2009). We thus still expected that the participants numbers were sufficient for our study. When comparing the results from the Anagram Task with the difference score in the Latin Buoyancy Task, we found that the participants who scored buoyantly on the Anagram Task (positive scores) generally also scored buoyantly on the Latin Buoyancy Task. For, even when there was a negative difference score between the pre- and post-setback task, this was less than the average decline of -2.19. Claudia and Quintina were notable. In the Anagram Task, they were non-buoyant, but in the Latin Buoyancy task they did seem more than averagely buoyant. We were aware from the results from Chapter 1, that the Anagram Task was not a perfect predictor for buoyancy in the Latin Buoyancy Task. This might explain Claudia's scores. However, we return to Quintina in further detail in Chapters 5 and 6 and propose an explanation for her unexpected buoyancy in the Latin Buoyancy Task.

To minimize problems arising from purposeful sampling, Palinkas et al. (2015) and Kemper, Stringfield & Teddlie (2003) provided recommendations, which we followed in the implementation of the extreme sampling. Extreme sampling is specifically aimed at identifying differences within participants and providing a large variance (Palinkas et al., 2015). Some might argue that the extreme sampling of the participants led to distorted results, as it only focused on participants with extreme levels of (non)-buoyancy (Patton, 1990; Patton, 2002). Looking at the results as presented in Table 2.3, we do observe that there are large differences between the participants: the range of the individual results is large and the standard deviations tend to be quite high. However, the aim of the dissertation is to establish how the setback task affects the participants' use of cognitive, metacognitive and affective strategies. We thus do not aim to generalize about academic buoyancy, nor to generalize about academic buoyancy in gifted students. Therefore, distortion is not an actual risk. Therefore, the results are not at risk of being distorted.

Operationalising giftedness was not straightforward, due to a lack of a broadly accepted definition of the term. We opted to only include think aloud participants who attended a *Begaafdheidsprofielschool*. Moreover, we assumed that because our participants were Latin students, they would belong to the highest cognitive potentials in the Netherlands. By making use of this method, we were able to avoid being purely reliant on unreliable IQ-tests. Interestingly, most of our participants had actually undergone an IQ test. As we focused on

Latin students, we automatically lost out on a group of gifted participants who for whatever reason did not have Latin in their curriculum. However, the gifted population is particularly varied (Reis, Sullivan & Renzulli, 2021). An exhaustive representation was not the aim of our study, but a diverse group of participants was. We might not have included a fully exhaustive range of gifted participants, but our participants do represent the group of students that at least Latin teachers see in their classroom.

Branch (2000) raised the possible issue of cognitive overload hindering participants' ability to think aloud. As we expected due to either the high cognitive abilities of our participants (Wade, 1990), the lack of time pressure (de Jong, 2010) or the recognisability of the translation task (Schnotz & Kurschner, 2007), our participants generally seemed to have little difficulty with thinking aloud. For example, in contrast to Wilson (1994) we did not have any deficient verbal transcripts. There were some differences between the participants regarding their ability to think aloud, which was also observed by Young (2005) in a think aloud study.

Overall, Diana in particular stood out for struggling with the thinking aloud, but despite cognitive load maybe being the reason, we propose that, in the case of gifted students, it more likely to be related to something else. Compared to the rest of the participants, Diana struggled somewhat with continuously thinking aloud and needed quite regular prompting to do so. Possibly her first emoji choice (tired) goes some way to explaining this. Something that might have hindered Diana besides overload/tiredness was trust and confidence levels. Gifted students are prone to having trust issues (Fornia & Frame, 2001; Winsor & Mueller, 2020) and problems related to fear of failure (Dai, 2000). The setting with the researcher was aimed at being as relaxed as possible. We emphasized that it was not an exam, no information would be shared with their teacher and we followed Ericsson and Simon's advice to include a warm-up task so that the participants got used to thinking aloud before the Latin Buoyancy Task begun. This seems to have indeed helped the participants relax into the thinking aloud, but possibly this was not enough for Diana to have been put fully at ease. Not until the final emoji, does she report a particularly outspoken positive emotion (happy), after reporting feeling curious and interested which are more directed at the task than the participants themselves. This leads us to wonder whether for particularly uncertain gifted students, the presence of an unknown researcher inadvertently led to the thinking aloud process becoming more difficult.

Within our mixed-methods design, we also made use of other methods to collect data besides the concurrent thinking aloud. The aim of this was to deepen our understanding of the results and academic buoyancy. Particularly the combination of the concurrent and retrospective thinking aloud seemed to enhance the data. On the one hand these two methods provided us with the opportunity to compare and contrast what the participants actually did with what they said they did. For example, in the retrospective thinking aloud, there were multiple utterances found that seemed to lean to desirable answers and did not actually represent what happened during the concurrent thinking aloud. Multiple participants approached the translation task linearly, sticking together word meanings from left to right. However, in the retrospective, these students said they translated the sentences by starting with the verb, a preferred translation strategy of many teachers (e.g. Kitchell, 2000; Kuhlmann, 2015). One participant even said that they did it 'how they have to translate according to their teacher'.[1] There were thus differences between the concurrent thinking aloud and the retrospective thinking aloud.

On the other hand, the combination of the concurrent and retrospective thinking aloud provided an opportunity to clarify why they made certain choices and what they meant. An example of this was regarding why despite a diminished motivation, Claudia still finished the task:

> Claudia: "Straightaway I felt that I understood it."
> Researcher: "And when you have that feeling, what happens to you?"
> Claudia: "Well, I am more motivated to finish the sentence and I feel better. […]"
> Researcher: "But if you're saying that you are less motivated by the second sentences, why did you still finish them?"
> Claudia: "Because that is what is expected of me. I just do that, if it is good or bad. Also, when I do homework. Sometimes I have no idea what I am doing, but I just write something down."[2]

Including both sorts of thinking aloud was, thus, valuable to the data collection.

---

[1] *Zoals we moeten vertalen van mijn docent.*
[2] Claudia: *Ik had meteen het gevoel dat ik het begreep.*
Researcher: *En als je dat gevoel hebt, wat gebeurt er dan met jou?*
Claudia: *Nou dan ben ik meer gemotiveerd om de zin af te maken en voel ik me er beter bij. […]*
Researcher: *Maar als je zegt dat je minder gemotiveerd was bij de tweede zinnen, waarom heb je ze toch afgemaakt?*
Claudia: *Omdat dat is wat er van mij wordt verwacht. Dan doe ik dat gewoon, of het nou goed of slecht is. Ook als ik huiswerk maak. Soms heb ik geen idee wat ik doe, maar schrijf ik maar wat op.*

## 5 CONCLUSION

In this chapter we presented how our study's data was collected using a mixed-methods design. Concurrent and retrospective thinking aloud formed the heart of the data collection regarding academic buoyancy.

We selected six buoyant and ten non-buoyant participants via extreme sampling. We also operationalised giftedness by selecting participants who attended a *Begaafdheidsprofielschool* and from students with Latin in their curriculum. This meant we were only screening a select population of gifted students. However, we trust that this method of defining gifted students was more reliable than only making use of IQ scores. When studying gifted students, particularly in a culture where giftedness is portrayed as more than a mere IQ-score, researchers must find other ways of interpreting what makes a participant 'gifted'. In our opinion, paying more attention to what makes a participant 'gifted' would enhance the gifted research field.

# CHAPTER 3

## TAPPING INTO TRANSLATION PROCESSES II: THE CODING SCHEMES

Where chapter 2 focused on how the data was collected, the current chapter aims to first explain how the think aloud recordings were transcribed into verbal transcripts and how these transcripts were subsequently analysed. Particular attention is given to the development of our main coding scheme and then to the development of a shortened coding scheme which was derived from the main coding scheme. These coding schemes were developed in multiple phases, which are each presented in this chapter. The purpose of the main coding scheme was to code the content of the verbal transcripts, focusing on the level of an individual utterance. This led to a list of 32 codes of cognitive strategies, metacognitive strategies and affective strategies. This coding scheme was used for the quantitative analysis of Chapter 4. For the qualitative analyses of particularly Chapter 5, the main coding scheme was too exhaustive. Based on the findings from Chapter 4, we opted to adapt the coding scheme to deepen our understanding of particularly the affective strategies and their relation to the cognitive and metacognitive strategies.

As demonstrated in the previous chapter, thinking aloud was integral to our data-collection. The data which this yielded is analysed in Chapters 4-6. However, to do so, a coding scheme was necessary. The development of the coding schemes we used to analyse our data and the coding itself are the central topics in the current chapter. However, before we can turn to those topics, we must first focus on how the think aloud recordings were turned into verbal transcriptions.

For, Ericsson & Simon (1993) stated that researchers should create verbal transcriptions to ensure the validity and integrity of the analysis of data collected via thinking aloud. Transcribing the recordings has become common practice in think aloud studies (e.g. Ericsson & Simon, 1993; van Someren, Barnard & Sandberg, 1994; Austin & Delaney, 1998; Young, 2005; Leighton, 2017). Often, however, researchers omit the choices regarding how the transcriptions came to be from their reports. As the current chapter aims to describe how we developed the coding scheme used to analyse the verbal transcripts and explain the coding process, we first present how the verbal transcripts themselves came to be readied for coding.

Keys (2000) put forward that the recordings should be fully transcribed, whereas Sun (2011) believed that full transcription of the recordings is unnecessary as information from the recordings is always lost, however thoroughly the researcher tries to transcribe everything. Sun thus proposed to only transcribe utterances directly related to the research topic. As our research question pertained to how the participants' translation process was affected by the setback task, we fully transcribed the recordings as advised by Keys (2000). The verbal transcripts were then segmented and coded for analysis using a coding scheme according to procedures described by Powers (2005).

One of the other decisions that must be made when creating verbal transcripts is who will perform the task of transcription (Tilley & Powick, 2002; Hennessy, Dennehy, Doherty & O'Donoghue, 2022). Researchers have previously indicated that transcription is an 'interpretive act' (e.g. Green, Franquiz & Dixon, 1997; Bird, 2005; Ten Have, 2007; Clark, Cuthbert, Lewis-Fernández, Narrow & Reed, 2017). In other words, the action of transcribing recordings is always done through the interpretive understanding of the transcriber and in extreme forms can even affect the outcome of the study (Tilley, 2003a; 2003b). Bird (2005) described the benefits of the researcher transcribing, particularly, as was the case in our study, when the researcher had also been present at the data-collection: it leads to a deeper understanding of the data. Moreover, Bazeley (2013) advised not to outsource the transcription, for similar reasons. Thus, for this study, the researcher transcribed the recordings.

We transcribed the recordings without the use of voice recognition software. Despite the rapid technological advancements of AI and some researchers mentioning the advantages of AI-generated transcriptions (e.g. Tilley, 2003a), there are still many interpretations AI is unable to make, such as the meaning of tone (McMullin, 2023). Time saved on the transcription itself might be lost again as the generated transcriptions must be scrupulously checked by the researcher (Bokhove & Dowey, 2018). Moreover, there are still unresolved ethical issues surrounding the use of AI in research (Da Silva, 2021). We, therefore, created the transcriptions ourselves, making use of only the recordings and Microsoft Word.

Another decision that must be made when transcribing is whether to create transcriptions that were 'full verbatim' (as the participants said things, including for example incongruent sentences) or 'intelligent verbatim' (i.e. grammatically corrected or with repetitions removed) (Bucholtz, 2000; McMullin, 2023). We were studying the effect of a setback, thus emotions such as frustration would probably be present (Jonassen & Grabowski, 2012; D'Mello, 2013). As language is affected by emotion (Barrett, Lindquist & Gendron, 2007), we believed that the participants' affection would be better reflected in a full verbatim transcript. However, we did add interpunction. When a part of the recording was inaudible this was marked in the transcript as [...].

Finally, according to Lapadat (2000), researchers must also decide whether to add other information to the transcripts. Besides transcribing the participants' utterings, we also included references to other audible utterances of the participants such as sighing or laughing. We did so as we felt certain behaviour might help interpret the participants' emotions (e.g. Farley, Risko & Kingstone, 2013). We also included an asterisk when the tone of the utterance might be determinative for the coding. Adhering to these decisions, we transcribed our recordings and created 16 verbal transcripts.

Before we were able to code these transcripts, they needed to be segmented (Ericsson & Simon, 1993; Payne, 1994; Van Someren, Barnard & Sandberg, 1994; Keys, 2000). Segmenting is a first step in the coding process and entails splitting the verbal transcripts into smaller chunks (Payne, 1994; Keys, 2000). In our case, we defined a chunk as 'one or more utterances that indicated that the participant was employing the same thinking activity'. This definition meant that the segmentation was closely related to the coding: without labelling the thinking activity, we were defining when the participants changed their thinking activity. In practice, segmentation and coding cannot be seen as two completely independent activities. Therefore, the segmentation was performed by one researcher and checked by another.

Our definition also meant that the segments differed in length and sometimes could be long if the participant did not switch to a different type of thinking activity. Figure 3.1 illustrates how the verbal transcripts were divided into segments, based on thinking activities. In this example the participant said 'But I don't think that is right'. This thinking activity (which we later labelled with the code 'monitoring') and, therefore, also the segment came to an end when the participant switched to another type of thinking activity, in this case translating two words into Dutch (later labelled as 'using Latin knowledge'). This segment quickly ended when the participant encountered a problem. This is followed by a slightly longer segment, wherein the participant defined what was getting in their way of progress, as later indicated by the code 'problem defining'. This segment ended when the participant switched back to an activity we later coded as monitoring.

*Figure 3.1 Illustration of the segmentation of a verbal protocol (Ennius)*

| Segment | Utterance (Dutch) | Utterance (English) | Thinking Activity |
|---|---|---|---|
| 8 | Maar dat klopt volgens mij niet. | But I don't think that is right | Monitoring |
| SWITCH | | | |
| 9 | Hier zingt. | Here sings | Using Latin knowledge |
| SWITCH | | | |
| 10 | Ik weet niet wat ik met 'zonder' moet, want dan wordt dit de nominativus. Er is geen zin die ik kan bedenken waar dit in de accusativus is. | I don't know what I have to do with 'without', because then this becomes the nominative. I can't think of any sentence in which this is the accusative. | Problem defining |
| SWITCH | | | |
| 11 | Misschien klopt dat al niet? | Maybe that is wrong? | Monitoring |

In total, the 16 transcripts contained 47640 words that were divided into 3917 segments. Rufus' verbal transcript was divided into the fewest segments (151) and Marcus' contained the most segments (334). The average length of a segment was 12.16 words. Some segments encompassed only one word, and on the other side of the spectrum, some segments contained up to 134 words.

The segments of the verbal transcripts were now ready to be labelled, or in other words, coded. Coding is the process of systematically categorizing the segments in verbal protocols (e.g. Creswell, 2014; Saldaña, 2016; Saldaña, 2021). Coding relies on a so-called coding scheme (Silver & Lewins, 2014). The creation of a coding scheme goes through multiple stages prior to the final coding (Neuendorf, 2016; Saldaña, 2016). Each stage allows for further clarification and honing of the coding scheme until it is deemed sufficient for the final coding. We thus needed a coding scheme befitting our research question (Van Someren, Barnard & Sandberg, 1994).

The aim of this dissertation is to establish how a lack of academic buoyancy in gifted students is reflected in their task process. To determine task process, we were particularly interested in coding the verbal transcripts for strategies employed by the students whilst translating. O'Malley and Chamot (1990) observed, analyzed and categorized strategies that students employed during language related problem-solving and text comprehension. They distinguished three types of learning activities: cognitive, metacognitive and affective (c.f. Lui, 2010).

According to Vermunt (e.g. 1992; 1996) cognitive strategies are related to knowledge and understanding. Metacognitive activities, on the other hand, are used to control and monitor cognitive strategies. Affective strategies are related to the emotions that students feel whilst learning. All three types of learning strategies are simultaneously at work during for example a translation task. Vermunt summarized the different cognitive, metacognitive and affective activities students might engage in in Table 3.1 (Vermunt, 1996, referencing Vermunt, 1989; 1992).

*Table 3.1. Summary of learning activities according to Vermunt (1996, p. 26)*

| Types | Categories |
|---|---|
| Cognitive | Relating, structuring, analyzing, concretizing, applying, memorizing, critical processing, selecting. |
| Affective | Attributing, motivating, concentrating, judging oneself, appraising, exerting effort, generating emotions, expecting. |
| Regulative[1] | Orienting, planning, monitoring, testing, diagnosing, adjusting, evaluating, reflecting. |

*Note. [1] Regulative is a synonym for metacognitive used by Vermunt.*

Learning strategies are strategies that are performed with the goal of learning something new (Vermunt, 1996; Lui, 2010). In the case of translation, the thinking activities are not strictly learning activities that are aimed at learning something, but translation is a skill that is honed by practice. When translating, many different cognitive knowledge bases are activated at once (e.g. Göpferich, Jakobsen & Mees, 2009). Moreover, from previous Latin translation studies, it is known that metacognition is also activated (e.g. van Houdt, 2008; Luger, 2020) and there are clear indications that affection influences the translation process (Newland, 2016; Bartelds, 2021). Therefore, we assumed that the terms ‘cognitive’, ‘metacognitive’ and ‘affective’ strategies could be applied to our coding scheme for a translation task, despite translation as performed in this study not strictly being a learning activity.

The remainder of this chapter presents how we developed the coding schemes used for the analysis of Chapters 4-6. These coding schemes contained codes for different cognitive, metacognitive and affective activities. To create the coding scheme, we implemented multiple design cycles, which are presented and discussed in the current chapter alongside the coding process itself. We first fully present the development and implementation of the main coding scheme. This coding scheme aimed at coding the individual utterances on a detailed level. As becomes clear from the discussion of the coding scheme a second, less exhaustive, scheme was necessary for qualitative analysis. In the latter coding scheme, we also incorporated the data from the retrospective thinking aloud. How we derived this second coding scheme from the main scheme is presented following the discussion on the main coding scheme.

In Chapter 4, the segments themselves, their length and the actual content of the segments were leading to answer the research question. Thus, in Chapter 4, the main coding scheme was the most relevant and consequently we only made use of the main coding scheme. In Chapter 5, we focused on the setback task itself, and we expected emotions to be at play as the task progressed. Therefore, made use of the derived coding scheme in Chapter 5, as this provided more insights into affection. Finally, in Chapter 6, we make use of both coding schemes alongside each other for an in-depth case analysis.

# 1 METHOD

## *1.1 Participants*

The creation of the coding schemes was driven by the verbal transcripts of the 16 participants. For more information on the selection of the participants and the participants themselves, see Chapter 2.

## *1.2 Materials*

The segmented verbal transcripts were the materials at the heart of this chapter. The creation of these transcripts and their segmentation have been presented in the Introduction of the current chapter.

## *1.3 Procedures*

In the coming sections, we first put forward the procedures which led to the creation of our main coding scheme. The creation of the main scheme and the subsequent coding took place in five different phases. We present each of these phases individually.

*Phase 1: Literature-driven*

The first stage in creating our coding scheme was to gather codes relating to cognitive, metacognitive and affective thinking activities from existing coding schemes. Our search for codes relating to cognitive, metacognitive and affective thinking activities focused on two different fields of thinking aloud research akin to our study: [1] Latin translation studies and [2] metacognition studies. In the following paragraphs, we first present the codes we derived from previous Latin translation studies that made use of thinking aloud. Then we explain how we included codes from existing metacognition research.

To determine which codes to include in our first version of the coding scheme, we consulted four studies that included thinking aloud to study Latin translation processes. These studies were Eikeboom (1970), Sarkissian (2008), Karten (2015) and Florian (2015). Many of the codes included in these studies were of a cognitive nature and closely related to Latin such as 'using background knowledge', 'focusing on verb tense', or 'focusing on word meaning'. As we were not specifically interested in the specific grammatical strategies used by the participants, but more generally in their process, we bundled the codes related to syntactical or morphological analysis together. Besides the code *Using syntactical or morphological knowledge*, we included four other cognitive codes derived from existing studies such as *Using the notes*. An overview of these cognitive codes can be found in Table 3.2 at the end of this section.

Table 3.2 also contains the metacognitive codes that we derived from existing coding schemes found in studies by Veenman (e.g. Veenman, 2011; Meijer, Veenman & Van Hout-Wolters, 2006). We, for example, included the metacognitive code *activating prior knowledge*. This code differed from the cognitive Using worldly knowledge in that activating prior knowledge was more of an associative orienting nature and does not directly lead to learning or solving the task (c.f. Vermunt, 1992; 1996). If a participant, without further inferring referred to general knowledge related to the sentences, this would fall under the code Activating prior knowledge. However, using the same example, if that participant then used that knowledge to determine if Caesar or another person in the text was commanding the soldiers, it would be coded as Using worldly knowledge, as it would then directly be used to come to a translation (c.f. Vermunt, idem).

The quotes 'Caesar is that general.'[1] and 'Because he is the god of the sea, I think it might be ships.'[2] further illustrate the difference between '*activating prior knowledge*' (no inference) and '*using worldly knowledge*' (inference). As becomes clear from the examples, in the case of activating prior knowledge, there is no explicit inference: the student only stated that Caesar was a general. On the other hand, in the case of the using worldly knowledge example, the student not only says that he (Neptune) is the god of the sea, but also uses that knowledge to infer the meaning of another word.

We also derived the codes *task orientation, monitoring* and *evaluating* from the think aloud studies by Veenman. *Task orientation* relates to activities that a student performs before actually beginning to perform the task that are aimed at preparing themselves for the task at hand (e.g. Pressley & Afflerbach, 1995; Veenman, Wilhelm & Beishuizen, 2004). The codes *monitoring* and *evaluating* both relate to checking their task progress: *monitoring* is done whilst performing the activity, for example 'am I doing what I should be doing?', whereas *evaluating* takes place after the activity, for example 'did I do the right thing?' (e.g., Veenman & Van Hout-Wolters, 2006; Veenman, 2011). Examples from the verbal transcripts of these metacognitive codes can be found in quotes 3.1-3.

---

[1] *Caesar is die generaal.*

[2] *Omdat hij de god van de zee is denk ik dat het misschien schepen is.*

Quote 3.1 Task orientation:
'Let's see what we have got.'[1]

Quote 3.2 Monitoring:
'I am doing this, but I do not think that that is right.'[2]

Quote 3.3 Evaluating:
'It went really badly.' [3]

Veenman's studies also incorporated 'goal orientation' as a metacognitive activity. As translating Latin texts into Dutch is a task the participants were used to performing regularly, we did not expect the participants to engage in explicit goal orientation. After all, when a student encounters similar tasks frequently, as is the case with translating, their metacognitive understanding of the task improves (Lawanto, Minichiello, Uziak & Febrian, 2019). Moreover, if their understanding of the task and thus their goal orientation had become automated, they would not explicate this (Schellings, 2011). We therefore did not include 'goal orientation' as one of our codes. However, we did expect the participants to encounter smaller specific problems whilst translating. An example of such an utterance found in our verbal transcripts was 'I cannot find the verb'[4]. For that reason, instead of goal defining, we included the code *problem defining*.

In Meijer, Veenman and Van Hout-Wolters' (2006) coding scheme, we also came across the code 'giving up'. We also included this code, as we expected participants to give up during the setback task due to them becoming frustrated. We, therefore, included this as an 'affective strategy'. With *giving up* we had derived 12 codes from existing think aloud literature to form the basis of our coding scheme as found in Table 3.2.

---

[1] *Even kijken wat hebben we.*
[2] *Ik doe het zo, maar denk niet dat het klopt.*
[3] *Het ging zo slecht.*
[4] *Ik kan de pv niet vinden.*

*Table 3.2 Codes derived from pre-existing coding schemes*

| Strategy type | Code | Definition |
|---|---|---|
| Cognitive | CSMSK | Using syntactical or morphological knowledge |
| | CSSK | Using semantic knowledge |
| | CSL | Using worldly logic |
| | RL | Reading Latin |
| | UN | Using the notes |
| Metacognitive | TO | Orientating on the task |
| | APK | Activating prior knowledge |
| | DP | Defining what the problem is |
| | MON | Monitoring their process |
| | EV | Evaluating their outcome |
| | U | Being uncertain about outcome |
| Affective | GU | Giving up on the task |

*Phase 2: Data-driven revision of the coding scheme*

Following the example of studies such as Meijer, Veenman and Van Hout-Wolters (2006) and Saldaña (2016), the second phase of developing our coding scheme was data-driven. Two researchers looked for utterances that were not yet covered by the codes selected from the existing literature. The researchers made suggestions for codes. These suggestions were compared and discussed during a meeting between the researchers. Eventually, we added one new strategy type, 'Other' and 18 codes to the list, some of which were to become subcategories of existing codes. These are presented in Table 3.3. In the coming paragraphs, we present the additions and alterations to the coding scheme first for the cognitive strategies, then for the metacognitive strategies and, finally, for the affective strategies.

We added four codes to the cognitive strategy category. The first of these were closely related to the producing of a Dutch translation. On the one hand, we added the code *formulating,* which pertained to utterances in which the participant was verbalizing based on what they knew of the Latin sentence to form a Dutch sentence. A quote that exemplifies *formulating* is "The horse with a spear pierced is."[1]. We similarly added *paraphrasing,* which referred to participants recapitulating the meaning of their translation in their own words. For example, "so, a spear has pierced the horse."[2] was coded as *paraphrasing*,

[1] *Het paard met een speer doorboord is.*
[2] *Dus een speer heeft een paard doorboord.*

because it was not a direct translation of the sentence, but phrased in the participant's own words.

In addition, we also included the codes *just* and *guessing*. In both cases these codes entailed the participants not acting upon considered deliberations. They did differ, though, regarding the extent of their lack of deliberation. In the transcripts, we found frequent instances when the participants admitted to just doing something, for example *'just writing something down as a translation'*[1] instead of basing their translation on morpho/syntactical analysis. Utterances of this type were coded as *just*. The code *guessing* was used in instances when participants gave themselves multiple options and randomly choose one of these. For example, *armis* could be a dative or ablative and some participants expressed that they randomly picked the case and, thus, guessed. The coding scheme now included a total of nine cognitive codes.

In the case of the metacognitive strategies, we further specified the code *evaluation* by splitting it into *positive evaluation* (e.g. 'It is correct up to here'[2]) and *negative evaluation* ('It went so badly'[3]). We also added four new metacognitive codes. First, we found examples of participants stating that under normal circumstances they would ask the teacher for help (e.g. 'I would now ask the teacher'[4]). Thus, we added *asking teacher for help* as a code. In our study they did not actually ask for help, merely mentioned it as what they would have done. Secondly, there were instances when participants referred to strategies they might employ to solve a problem, but did not actually employ. As this attested to a certain awareness, despite the participants not actually using it, we also included *naming a strategy they could but do not use* as a code. An example of this can befound in the quote 'I could check it'[5] We also included the code *postponing* for postponing solving an encountered problem (e.g. 'I am just skipping this one'[6]). Finally, some participants attested to doing something because their teacher had previously given that advice, for example, to move on after a certain amount of time spent on trying to solve a problem ('According to my teacher, I should now continue'[7]). We thus added the code *learned or instructed strategy*. The refinements and additions led to our new version of the coding scheme containing 11 metacognitive codes.

---

[1] *Gewoon iets opschrijven als vertaling.*
[2] *Het klopt tot hier.*
[3] *Het ging zo slecht.*
[4] *Ik zou het nu aan de docent vragen.*
[5] *Ik zou het kunnen controleren.*
[6] *Ik sla deze even over.*
[7] *Van mijn docent moet ik nu verder.*

In the case of the affective codes, we began by further specifying the code *giving up*. For, the transcripts gave indications of some participants seemed to give up because they wanted to stop ('I've had enough'[1]), whilst others seemed to be giving up out of necessity but not because they wanted to ('It's not correct, but I cannot make any more of it'[2]). These reasons for stopping differed in that the first only made a reference to wanting to stop, regardless of the quality of their task completion. The second quotation did reference the quality: in this case, despite being aware that the translation was not correct, the participant said they would stop, because they felt there was nothing else they could do to improve it. We thus distinguished between outright *quitting* and *consciously deciding to stop*. In contrast to stopping, we also included the code *persevering*. A quotation that exemplifies this code is 'try again'[3]. We categorized this as *persevering* as the participant encouraged himself to continue to look for a solution to a problem he had up until now been unable to solve.

We further incorporated codes related to judgment in the coding scheme. These codes pertained to utterances that indicated that the participant was expressing an opinion on the task as a whole. We distinguished between two types of judgement. On the one hand, the verbal transcripts included phrases such as 'It was good because it was short'[4]. In this case the participant was expressing a positive opinion towards the task. On the other hand, there were also examples of negative opinions, for instance, 'it is difficult'[5]. We thus in included the codes *negative judgment* and *positive judgment.*

Finally, we added the category 'Other', which was to include a code for any procedural questions, such as 'May I write the paradigms down?'[6] posed by the participants. After this addition, we were left with a total of 28 codes. All additions to the scheme are presented in Table 3.3.

---

[1] *Ik ben er klaar mee.*
[2] *Het klopt niet, maar ik kan er niet meer van maken.*
[3] *Nog een keer proberen.*
[4] *Het was fijn omdat het kort was.*
[5] *Het is moeilijk.*
[6] *Mag ik de rijtjes opschrijven?*

*Table 3.3 The data-driven additional codes*

| Strategy type | Code | Definition |
|---|---|---|
| Cognitive | CSJ | Just |
| | CSG | Guessing |
| | CSP | Postponing |
| | F | Formulating their translation |
| Metacognitive | AT | Hypothetically reaching out to teacher for help |
| | ST | Naming a strategy, they could but do not use |
| | P | Paraphrasing their translation |
| | LI | Referring a learned or instructed strategy |
| | EP | Evaluating their task outcome positively |
| | EN | Evaluating their task outcome negatively |
| Affective | GUS | Giving up on the task: quitting |
| | GUR | Giving up on the task: *consciously deciding to stop* |
| | Gr | Explicitly referring to grit or perseverance |
| | JP | Judging positively |
| | JN | Judging negatively |
| | MOTN | Not being motivated to complete the task |
| Other | PQ | Procedural questions |

*Phase 3: Testing and evaluating the coding scheme*

Following the recommendations of Van Someren, Barnard and Sandberg (1994) and Bird (2005), two of the study's researchers used the current coding scheme to fully code two verbal transcripts in Atlas.ti. In a final coding session, each segment would have to be labelled using only one code. However, in the current phase, we were on the look-out for boundary cases, and would, therefore, in the case of doubt, assign multiple codes. Then the researchers met up for a consultation session. During this consultation, we compared and discussed our experiences with using the coding scheme. We exchanged examples of segments we found difficult to code and codes we felt were still particularly broad and, thus, actually little indicative of the participants' translation processes.

The difficulties particularly arose around the codes related to evaluation and judging. It was discussed that evaluation pertained to utterances in which the participants were reflecting upon something directly related to the problem at hand, whereas judgment related to more general reflections on for example the task as a whole. The exchange led to two further alterations to the coding scheme.

First, we split the code for *evaluation* further. For, participants were not only evaluating their translation outcome but also other things. We accommodated the different types of evaluation by creating separate codes for whether the participants were evaluating their translation (e.g. I think this is correct[1]), their abilities (e.g. I am not good at translating[2]), or their knowledge (e.g. I don't know the case endings well enough[3]). The second alteration was to also split the codes relating to *judgment*. It seemed necessary to distinguish between the instances when the participants were judging themselves personally (I am just bad at translating[4]) from when they were judging the task (I dislike translating[5]). These additions to the coding scheme are found in Table 3.4.

The two researchers fully re-coded both verbal transcripts alongside coding a third transcript using the adjusted coding scheme. This led to a coding scheme including 33 codes (for a complete overview of the coding scheme, see the Results section). A second consultation did not lead to further changes and thus the coding scheme was deemed ready for pre-coding of all the remaining transcripts.

*Table 3.4 Further altered codes*

| Strategy type | Code | Definition |
|---|---|---|
| Metacognitive | AP | Evaluating their abilities positively |
| | AN | Evaluating their abilities negatively |
| | KP | Evaluating their knowledge positively |
| | KN | Evaluating their knowledge negatively |
| | SP | Judging themselves as a person positively |
| | SN | Judging themselves as person negatively |
| | TP | Judging the task positively |
| | TN | Judging the task negatively |

*Phase 4: Pre-coding and evaluation*

In the fourth phase, the two researchers coded 10% of each protocol. The aim of this round of coding was to ascertain whether the intercoder reliability was sufficient enough when using the current coding scheme. Of each transcript, a random 10% was selected. We opted to not code the first 10% of each transcript

[1] *Ik denk dat dit klopt.*
[2] *Ik ben niet goed in vertalen.*
[3] *Ik ken de rijtjes niet goed genoeg.*
[4] *Ik ben gewoon slecht in vertalen.*
[5] *Ik vind vertalen stom.*

to include a larger variety of codes that we might come across and thus safeguarding the validity of the reliability analysis for transcripts. Thus, the sample included parts of the transcripts of the pre-setback, setback, post-setback task and retrospective thinking aloud. Using this sample, a sufficient level of agreement between the coders ($\kappa$ = .842) was found (Fleiss, 1981; Lombard, Snyder-Duch & Bracken, 2017).

*Phase 5: Coding*

Using the final version of the coding scheme the remaining portions of the protocols were then coded independently by one of the researchers. At this moment in time, the researchers as yet had no knowledge of the participants' mindset preference and frustration tolerance. The data from the questionnaires would be addressed after the coding. As the coders were both experienced in marking Latin translation, they would be somewhat aware of the students' translation proficiency as they coded.

## 2 RESULTS

As presented in the previous section, we developed an exhaustive coding scheme, which aimed at representing the content of the participants' utterances. In this section we provide our main coding scheme in Table 3.5. We do so not only by including the 32 codes, but also by defining each code and giving examples of segments that might be coded with that code.

*Table 3.5 Main coding scheme: Related to the content of the utterances*

| Code | Content | Example English | Dutch |
|---|---|---|---|
| | | *Cognitive* | |
| CSSMK | The use of syntactical or morphological knowledge | It is an ablative | *Het is een ablativus* |
| CSSK | The use of semantic knowledge | *Equus* means horse | *Equus betekent paard* |
| CSL | The use of worldly logic | Because he is the god of the sea, I think it might be 'ships' | *Omdat hij de god van de zee is denk ik dat het misschien schepen is* |
| CSJ | Choosing the use of 'just' any type of knowledge | I am just going to write something down | *Ik ga maar gewoon wat opschrijven* |
| CSG | Choosing a strategy by guessing | I will have to guess | *Dan moet ik maar gokken* |
| RL | Reading Latin aloud | *Equus pilo traiectus est* | *Equus pilo traiectus est* |
| UN | The use of the notes | In the word list it says … | *In de woordenlijst staat…* |
| F | Formulating a translation | The horse with a spear pierced is | *Het paard met een speer doorboord is* |
| CP | Paraphrasing their translation | So, a spear has pierced the horse | *Dus een speer heeft een paard doorboord* |
| | | *Metacognitive* | |
| TO | Task orientation | Let's see what we have got | *Even kijken wat hebben we* |
| APK | Activation of prior knowledge | Caesar is that general | *Caesar is die general* |
| DP | Defining an encountered problem | I cannot find the verb | *Ik kan geen pv vinden* |
| MON | Monitoring their task process | I am doing this, but I do not think that that is right | *Ik doe het zo, maar denk niet dat het goed is* |
| AP | Evaluating their abilities positively | I can do it when I have prepared properly | *Ik kan het als ik goed heb voorbereid* |
| AN | Evaluating their abilities negatively | I am bad at translating | *Ik kan niet goed vertalen* |
| KP | Evaluating their knowledge positively | I know this for sure | *Ik weet dit zeker* |

| | | | |
|---|---|---|---|
| KN | Evaluating their knowledge negatively | I do not know the paradigms | *Ik ken de rijtjes niet goed* |
| SP | Evaluating their self or 'being' positively | I am a trouper | *Ik ben een doorzetter* |
| SN | Evaluating their self or 'being' negatively | I am uncertain if it is normal that I do not know this | *Ik weet niet of het normaal is dat ik het niet weet* |
| TP | Evaluating their task outcome positively | It is correct up to here | *Tot hier klopt het* |
| TN | Evaluating their task outcome negatively | It went really badly | *Het ging zo slecht* |
| U | Evaluating their task outcome uncertainly | I am not sure if it is correct | *Ik weet niet of het goed is* |
| AT | Hypothetically asking a teacher for help | I would now ask the teacher | *Ik zou nu de docent vragen* |
| ST | Formulating a strategy, they could but do not use | I could check it | *Ik zou het kunnen nakijken* |
| LI | Employing a learned or instructed strategy | According to my teacher I should now continue | *Van mijn docent moet ik nu doorgaan* |
| P | Postponing the problem at hand | I am just skipping this one | *Ik sla deze even over* |
| | | *Affective* | |
| GUS | Giving up: stopping | I am done with it | *Ik ben er klaar mee* |
| GUR | Giving up: resignation | It is wrong, but I have done my best | *Het klopt niet, maar, ik heb mijn best gedaan* |
| GR | Referring to persevering | Try again | *Nog een keer proberen* |
| JP | Judging the task as a whole positively | It was good, because it was short | *Het was fijn omdat het kort was* |
| JN | Judging the task as a whole negatively | It is difficult | *Het is moeilijk* |
| | | *Other* | |
| PQ | Procedural Questions | May I write the paradigms down? | *Mag ik de rijtjes opschrijven?* |

## 3 DISCUSSION

In the current section we discuss the main coding scheme and the coding process. However, first, we reflect on our choice to create and code the transcriptions ourselves. We did so on the advice found in studies by for example Tilley, 2003a, Bird, 2005 and Bazeley, 2013. Looking back on the process, despite it being a time costly activity (c.f. Charters, 2003), we did experience a particular benefit of transcribing and coding ourselves. The frequent close reading and discussing of the codes led to us becoming well-acquainted with the verbal transcripts (c.f. Bird, 2005; Bazeley, 2013).

The many codes made it possible for us to create a detailed overview of the different types of utterances the participants demonstrated. Using the first coding scheme, we, for example, saw similar cognitive activities as found in other think aloud studies pertaining to Latin translation (Eikeboom, 1970; Sarkissian, 2008; Karten, 2015; Florian, 2015; Boyd, 2018). Ours differed in that we did not split out every different type of syntactical or morphological activity into separate categories. The value of our coding scheme is that it is sufficiently detailed and builds a picture of the translation processes of our third-year Latin students, without it being overly focused on the different cognitive processes.

The inclusion of the retrospective thinking aloud provided us with a new insight: affective strategies seemed to differ depending on the learner variables during the setback. However, in the current coding scheme, affection remained somewhat underexposed with only very few examples of affective codes being found. This might be explained by Pickard's (2003) conclusion that emotions are more difficult to share than other thoughts. Another explanation might be related to affective judgments mainly taking place during reflection (Zimmerman, 2000; van de Velde, et al., 2015), and, therefore, more likely to be present during the retrospective thinking aloud compared to the concurrent thinking aloud.

Also, the retrospective thinking aloud led us to observe that there was sometimes a discrepancy between what the participants literally said during the concurrent thinking aloud and what they seemed to mean. For example, the utterance 'I'll just write something down' (coded with the cognitive code *just*) might actually be an indication of the participant giving up or avoiding the problem depending on the context. In such cases, an expression in the form of a cognitive or metacognitive utterance might actually be the result of an affective intention. We demonstrate this difference on the basis of two quotes from the verbal transcripts:

> 'Let's see, battle. Uhm *continebat* is not, oh wait, here it says 'to keep away'. Uhm, uh, keep away battle keep away Caesar. Uhm huh? It is logical that it would be something like 'Caesar keeps is name away from the battle', but I do not think that sounds like something Caesar would do. **Well if you do not know, just write it down right**. I hope that I can persevere. So, Caesar, how do I need to translate that again? Oh bummer. Uh, his men, that is then keep away with a t. So that is just o, s, t, no, bat. Oh god, what does that mean again? Uhm it was something like, no idea. I know that it is something, just no idea what. Uhm. It is something with, no, yes, imperative. No, I would not know what it is exactly. Well, uhm, *suos*, his name, but that is the object. That is probably genitive, it is somebody's? It is his name, and it says 'of' so that will be the case. So, the battle is probably the object. Uhm and keep away is probably the personal... but there was still something wrong with it. Uhm, yeah, uh, okay. Caesar, oh wait. Okay, we will get around to that later. Caesar keeps his name away from the battle. Finally.'
>
> - Marcus[1]

> He gives. He gives us maybe. He, what am I supposed to make of this? **If you do not know it, just write down a sentence that sounds logical**, then you are rid of it. Whatever. I will just write down 'He maybe gives us love and forgiveness.'
>
> - Octavia[2]

The passages from Figure 3.6 make clear that despite similar literal content, the underlying strategy differs: Octavia 'just writes something' to be rid of the task and does not continue after doing so. This use of 'just writing something down' thus, seemed more of a way to completing the task quickly than translating it

---

[1] *Even kijken strijd. Uhm continebat staat er niet o wacht staat hier weghouden. Uhm uh weghouden strijd weghouden Caesar. Uhm huh? Het is logisch dat er dan zoiets zou staan als Caesar houdt zijn naam weg bij de strijd, maar dat vind ik niet heel erg iets voor Caesar om te doen. Nou als je het niet weet dan toch maar gewoon iets opschrijven he. Ik hoop dat ik het volhoud. Dus. Caesar, hoe moet ik dat ook alweer vertalen? O ja kut. Uh zijn mannen, dat is dan uh weghouden, met een t. Dus dat is gewoon o s t, nee bat. Oh god wat is dat ook alweer? Uhm ja dat was iets met geen idee, ik weet dat het iets is, maar welke het is geen idee. Uhm. Het is iets met, nee, o ja imperativus. Nee, wat het precies is zou ik niet weten. Nou uh suos, zijn naam, maar dat is lijdend voorwerp. Dat is waarschijnlijk genitivus, het is van iemand? Het is zijn naam en er staat van, dus het zal wel. Dus dan is strijd waarschijnlijk het lijdend voorwerp. Uhm en weghouden is dan waarschijnlijk de persoonlijk. Maar daar was nog iets mis mee. Uhm ja. Uh. Ja goed. Caesar oh wacht. Oké zien we zo wel. Caesar houdt zijn naam weg van de strijd. Enfin.*

[2] *Hij geeft. Hij geeft ons misschien. Hij, wat moet ik hier nou van maken? Als je het niet weet maak je gewoon een logische zin en schrijf je dat op, dan ben je ervan af. Het zal wel. Ik schrijf gewoon op Hij geeft ons misschien liefde en vergiffenis.*

correctly, as indicated by ‘whatever’. On the other hand, Marcus similarly ‘just wrote something down’ and then used that to continue trying to come to a correct translation. At the same time, Marcus’ quotation displays a lot of underlying affect and Marcus even explicitly refers to persevering. As the fragment demonstrates, Marcus spent quite some further time after writing a possible translation down, improving said translation. In Marcus’ case, writing down seemed a helpful coping strategy and supported him persevering despite his affected state of mind. Yet, in the case of Octavia it seemed less helpful for coming to a correct translation and seems closer to avoiding trying to solve the problem at hand.

This distinction of helpful and unhelpful strategies was of particular importance for the analysis of the transcripts of the setback task as we expected this task to trigger the participants’ emotions the most due to its impossible nature (c.f. Klein, et al., 2019). We thus decided to create a second coding scheme, which aimed at providing more possibilities to code according to the affective context in which the utterance was made than the main coding scheme had yet provided. This scheme would then also allow us to focus more on the helpful or unhelpful strategies underlying the utterances.

Furthermore, where our interest in the role of affective strategies on cognitive and metacognitive activities grew during the qualitative analysis, the distinctions between the different cognitive activities proved to be less relevant to our research question for Chapter 5: the setback task did not much seem to affect exactly *which* cognitive activity the participants displayed, but more their cognitive activity in general. We, therefore, deemed it necessary to reduce the number of codes in the coding scheme, by creating a second scheme which was derived from the main scheme by clustering and re-evaluating the codes.

A second reason to reduce the number of codes, was related to the qualitative analysis we performed in Chapter 5. To analyse the translation processes of the participants during the setback task, we created visualisations. The exhaustive number of codes meant that these visualisations were too detailed and varied for a valid analysis. Not only the many cognitive codes were distortive to the visualisations, but also the splitting of the codes relating to judgment. This then provided a further reason to derive a second coding scheme from the main coding scheme. The derived scheme not only aimed at focussing more on the context in which the utterances where said, but also enabled a more meaningful qualitative analysis due to the reduced number of codes.

## CONTEXTUALIZING CODING SCHEME

As the participants and materials were the same as for the creation of the main coding scheme, we shall focus on the procedures which we followed to derive the second, shorter, coding scheme. The creation of this coding scheme took place after the quantitative analysis of Chapter 4. We deemed the creation of the scheme necessary to be able to include the retrospective thinking aloud in our analysis. As we had only looked at the relationship between the learner variables and the extent of buoyancy on the Latin Buoyancy Task, the coders still knew very little of the individual participants regarding their mindset preference and frustration tolerance. This then did not influence the coding. From this point onwards, this second coding scheme will be known as the 'Contextualizing coding scheme'. In the coming section we present the three phases in which we developed the coding scheme.

### 1 METHOD

*Phase 1: Literature review*

Again, we started with a literature review of Latin translation studies, this time specifically looking for observations regarding affective strategies. We generally found few references to affective learning activities. However, in a small-scale case study including only two participants the term 'avoidance' was introduced. (Newland, 2016). According to the APA Dictionary of Psychology (2015, p. 101), avoidance coping is a strategy that people can employ when stressed or emoted. People who are prone to avoidance coping avert their attention from the problem at hand or disengage from it. Seeing that the setback task might lead to stress, we included the code *avoidance* as a possible strategy underlying the utterances found in the verbal transcripts.

*Phase 2: Data-driven revision of the coding scheme*

After the literature review, we returned to the data to re-assess the need of certain (sub)codes on the one hand and on the other to discuss utterances with ambiguous codes, i.e., covering both helpful and unhelpful strategies. First, we concluded that codes concerning the use of separate types of cognitive knowledge were of less relevance for our current research question. Many of the utterances were directly aimed at coming to a translation. An example of such an expression is 'it is an ablative with in'[1]. The aim of this utterance could not be metacognitive of nature or indicative of their emotions. This and similar

[1] *Het is een ablativus met 'in'.*

utterances were of a cognitive nature, and we decided to make no further distinction between these utterances.

For the metacognitive codes, we included *monitoring* and *problem defining.* We defined *monitoring* as the participants checking their actions or outcome. We did not further distinguish between evaluating and monitoring here. This was a possibility to reduce the number of codes, because evaluation is highly dependent on monitoring (Pintrich, 2000) and thus the intention, i.e. to check, is the same. Moreover, previous studies have also determined them as one category (Cohors-Fresenborg & Kaune, 2007; Baumanns & Rott, 2022). *Problem defining* related to students defining what the problem was they were encountering.

We also included the code *stepping out*, for instance when the intention of the participant was to complete the task without emotions that instigated the wish to stop. An example of this difference can be found in the utterances 'I am now moving on to the next'[1] and 'Shall we move on to the next?'[2] expressed with a breaking voice and a worried facial expression. These quotes differ, in that the first is not driven by emotion or motivation loss, whereas the second is. This distinction led to the code *loss of motivation* being added to the list of affective codes.

Going through the verbal transcripts again, we noted that it was not always clear which emotions were exactly at play, as there were very few direct utterances in which emotions were explicitly mentioned besides the moments of filling in the emoji chart. However, we could distinguish between whether an utterance came from a positive emotion or a negative emotion without further labelling the emotion itself. We thus included the codes *negativity* and *positivity*. For example, 'o no not more' was coded as *judging the task negatively* using the main coding scheme. But, using the contextualizing coding scheme we coded it as *negativity*. We could not exactly determine whether this participant was disappointed or unmotivated by seeing that there were more sentences that needed to be translated, but in either case, it implied a negative feeling towards the task. On the other hand, saying 'I feel content', regarding their translation, comes from a positive emotion.

*Phase 3: coding*

As in the development of the main coding scheme, two of us re-coded the verbal transcripts of two of the verbal transcripts, using the contextualized coding

---

[1] *Ik ga nu naar de volgende.*

[2] *Zullen we naar de volgende?*

scheme. In doing so, we focused on the setback task and the retrospective thinking aloud.

Another consultation session was held. Particularly the codes *avoidance*, *positivity* and *negativity* were discussed, as these were the most difficult to code. Often the retrospective thinking aloud transcription provided help for the coding of such segments. The segments that were difficult to code were discussed and arguments were put forth to determine whether or not there were indications of emotion leading to the utterance. To determine *avoidance*, the context was of particular relevance. For example, the utterance 'maybe this is not a problem after all'[1], on its own, does not necessarily imply *avoidance*. However, this directly followed an observation that the verb and subject did not match in number. There was no reason to suppose that what only seconds earlier was deemed a problem suddenly was not a problem. Moreover, no attempt was made to solve the problem, the participant just got rid of the problem by saying that it was not a problem. The lack of attempt to solve the problem led it to being coded as *avoidance*.

Finally, all the segments pertaining to the setback task were re-coded by the two researchers using the second coding scheme. The full coding scheme can be found in the Results section. Using Cohen's kappa, a high level of intercoder reliability was found for the coding of the whole protocols over all the codes (.90).

## 2 RESULTS

As presented in the previous section, we developed a shortened coding scheme, derived from the main coding scheme. This contextualizing coding scheme aimed at representing the context and underlying strategies of the participants' utterances. In this section we provide our contextualizing coding scheme in Table 3.6. We do so not only by including the eight codes, but also by defining each code and giving examples of segments that might be coded with that code.

[1] *Misschien is het toch geen probleem.*

*Table 3.6 Derived coding scheme: Focusing on helpful and unhelpful metacognition and affect*

| Type | Code | Intention | Example English | Dutch |
|---|---|---|---|---|
| Cognitive | L | Using cognitive strategies to come to a translation | It is an ablative | *Het is een ablativus* |
| Metacognitive | M | Monitoring whether the process or outcome is as it should be | Oh wait | *Oh wacht,* |
| | D | Defining the problem at hand | I have no idea what I have to do with *cano* | *Ik heb geen idee wat ik met 'cano' aan moet* |
| | G | Progressing to a next sentence without being emotionally driven to do so | Okay, the next one | *Oké, de volgende* |
| Affective | A | Avoiding the problem at hand | Maybe it is not actually a problem[1] | *Misschien is het toch geen probleem*[1] |
| | ML | Progressing to a next sentence whilst being emotionally driven to do so | Shall we move on to the next set?[2] | *Zullen we naar de volgende set?*[2] |
| | N | Expressing a negative feeling | No, not more | *Nee, niet nog meer.* |
| | P | Expressing a positive feeling | I am feeling content | *Ik voel me tevreden* |

*Notes.* [1] *This is an example of avoidance as the problem defined immediately prior to this statement, was a valid problem as was the participant's reasoning that it was a problem. Instead of looking for a solution, the existence of a problem was ignored.* [2] *The additional recorded data indicated emotion through a broken voice.*

## 3 DISCUSSION

The number of codes in the exhaustive coding scheme impeded the creation of meaningful visualisations of the participants' translation processes. This was one of the motives to develop the shorter contextualized coding scheme. As demonstrated in Chapter 5, the contextualized coding scheme indeed led us to be able to create visualisations that we could then compare and contrast to each other. The combining of the cognitive codes from the main coding into one code *cognitive strategy*, did not hinder our understanding of the aspects of the translation processes in which we were interested.

The other motivation for the development of the contextualized coding scheme was to give us more room to ask ourselves why the participants said what they said and strengthen our qualitative analysis. This indeed proved the case and by using the contextualized coding scheme: it provided us with the ability to ask the 'why-question'. There was thus more focus on the helpful and unhelpful strategies employed by the participants whilst translating.

On a small scale, a study by Newland (2016) also asked the why-question for the freezing or avoidance he witnessed in two of his students. This descriptive case study of what two weak translators did during three Latin lessons, implies that they avoided certain activities due to a fear of failure. These students attributed their expected failure to a low perceived proficiency. In our re-coded verbal transcripts, we also came across multiple instances of *avoidance*, which would have been missed if we had only relied on the main coding scheme. In Chapter 5 we specifically look at the participants' translation process of the setback task and ask if differences in learner variables, such as Newland (2016) proposed for proficiency, affected how the participants dealt with the setback.

As we were unable to ask the participants 'why' during the coding, the retrospective thinking aloud was of particular importance when we were interpreting the 'why' necessary for coding with the second coding scheme. For, often during this phase of the thinking aloud data-collection, the participants would reflect on what they were feeling or the underlying reason for doing what they did during the translation tasks. This underlines the added value of using both concurrent and retrospective thinking aloud (see Chapter 2).

The recordings also helped us in ascertaining the 'why'. Despite having included references to for example intonation, sighs and laughter in our transcripts, it was not always sufficient information to confidently interpret the underlying meaning of the segment. This aligns with Sun's (2011) assertion that, however much information is included in verbal transcripts, they always remain incomplete and lacking in some regard. We thus ended up re-watching certain parts of the recordings to help us code some of the affective segments. We would, therefore, advise any researchers particularly coding affective strategies, to not only code from the verbal transcriptions but include re-watching the recordings to support their understanding and coding of the segments.

Validating the researchers' interpretation of 'why' participants did what they did could benefit from future research. Some existing think aloud studies ask their participants to check and approve the verbal transcripts before coding (e.g. Bird, 2005). Possibly one could also ask their participants to approve whether affective codes are a correct interpretation. However, the transcribing and coding could not be done fast enough to have the participant be able to reliably remember what they were thinking at the time (Wade, 1990; Kuusela & Paul, 2000; Young, 2005; see also Chapter 2). The use of AI for transcribing and/or coding might speed things up to some extent (Tilley, 2003a), but the interpretation of affective strategies is especially tricky for AI to interpret (McMullin, 2023). Thus, this does not seem a feasible possibility to validate the interpretation of the affection codes further.

Retrospective thinking aloud might provide a more feasible option for this. By having the researcher observe the progress of the concurrent think aloud and making notes of moments when it was not fully clear why the participant does something, it would be possible to structure the retrospective thinking aloud more. This would lead to more inclusion of the participants' answers to 'why?' helping the coders' interpretation. Future research pertaining to further validating the affective codes might also be found in the Emoji chart (see Chapter 1). Cross-referencing the affection codes with the participants' answers on the Emoji chart or a study to ascertain the extent of a relationship between the codes and the filled-in Emoji chart might provide more insights into the use of affective codes in thinking aloud studies. However, despite these possibilities using the Emoji chart, they were outside the scope of this dissertation. This is because our main aim was not to ascertain the best way in coding affective learning strategies, but to study academic buoyancy through a task-based measurement. Therefore, our current coding scheme did not aim to distinguish between emotions beyond whether they were driven by *positivity* or *negativity*.

Overall, the contextualized coding scheme might be more similar to what Latin teachers will observe during their lessons. For, even when their students are not translating whilst thinking aloud, these are the activities they are engaging in and discussing amongst themselves. By focusing less on exactly which cognitive activity is occurring, there is more room for guiding their metacognitive and affective strategies. These are also helpful learning activities and should be included in teaching alongside cognitive learning activities (Vermunt, 1996; Vermunt & Vermetten, 2004).

## 4 CONCLUSION

This chapter first presented how we created and segmented the verbal transcripts. However, the chapter's main focus lay in expounding how we created the two coding schemes used to analyse the verbal transcripts.

At first, we set out to create the main coding scheme. This coding scheme aimed at reflecting the content of the participants' utterances and therethrough study whether the participants' translation process changed throughout the task. However, we discovered that this coding scheme was not sufficient to answer a newly emerged research question regarding affection on the one hand, and to present the data of the setback task clearly and efficiently on the other hand. We, therefore, also created the contextualizing coding scheme that focused on distinguishing helpful and unhelpful metacognitive and affective strategies.

The difference between the two coding schemes proved to be mainly related to the participants' emotions. Participants were not prone to explicitly uttering 'affective' categories from the main coding scheme. However, the contextualizing coding scheme allowed more attention to the context of utterances, and, thus, more examples of the participants being emotionally affected by the task came to light. It is, thus, advisable when using a think aloud method to study emotional or affective phenomena to incorporate a way of discerning affection through other means than their direct utterances from the offset.

As this chapter concludes, we also arrive at a breakpoint in this dissertation. Where the first three chapters were of a methodological nature, the following chapters focus on the data-analysis. From this point onwards, establishing how the setback task affects gifted students' use of cognitive, metacognitive and affective strategies and how this relates to the students' extent of academic buoyancy takes centre stage.

PART II

# ACADEMIC BUOYANCY & LATIN TRANSLATION

# CHAPTER 4

# DIFFERENTIATING SETBACK EFFECTS
# A QUANTITATIVE ANALYSIS

In the previous chapters we explained how we collected our data to establish how a temporary academic setback affected gifted students' learning processes, in order to study academic buoyancy. In the current chapter we aimed to answer two questions: [1] To what extent are the participants translation quality and processes affected by the setback? and [2] To what extent is that the effect of the setback on the students' translation quality and processes moderated by their translation proficiency, mindset preference and frustration tolerance? To answer these questions, we adopted a quantitative analysis method of the verbal transcripts from the pre- and post-setback task. Using nested models, we compared and contrasted the accuracy of the students' translations prior to and after the setback. We also included three process variables within the comparison, to measure whether the setback affected the students' translation process. Finally, we included the learner variables in the analysis. We found that the translation accuracy was significantly affected by the setback. We also found that when a student was high proficient or inclined towards a fixed mindset, the translation accuracy declined significantly more than in the case of low proficient students or those inclined towards a growth mindset. Frustration tolerance was not found to be a clear moderator of the effect of the setback on the students' translation accuracy.

# 1 INTRODUCTION

'Oh no, more sentences! I thought I was rid of them. Oh, well, I might suddenly be really good at translating them.'[1]
- Phaedra

This quote above is from one of our verbal transcriptions of the Latin Buoyancy Task. At the time, Phaedra had just finished trying to translate our setback task and had been presented with the post-setback task. Such situations are not uncommon in education: tasks must be completed within a certain timespan and subsequent tasks can differ in difficulty. As Phaedra's quote indicates, a setback can affect a student's frame of mind. At first, Phaedra seemed disappointed that the Latin Buoyancy Task consisted of more sentences. But then suddenly, without any external intervention, Phaedra found something positive: this time around, translating might go better. Indeed, as Phaedra hoped, translating the post-setback task went rather well, despite having struggled much during the setback. The accuracy of Phaedra's translation bounced back to the level of translation from the pre-setback task.

In contrast to Phaedra, most of the other students who performed the Latin Buoyancy Task were unable to recover after the setback task. For example, 15-year-old Gaius reflected on the setback task: "[When I'm frustrated], I am less bothered about taking my time to read everything and to look properly."[2]. Gaius explained that the experience of the setback task had led to a change in translation process during the post-setback. The accuracy of Gaius's post-setback translation had deteriorated significantly compared to his score from the pre-setback task. This deterioration might be explained by a difference in Gaius's translation process.

Gaius and Phaedra differed with respect to their academic buoyancy. Martin and Marsh (2008; 2013) defined academic buoyancy as the ability to bounce back from everyday academic setbacks. As described in Chapter 1, our Latin Buoyancy Task was designed to mimic a daily academic setback that students might encounter. As Phaedra and Gaius demonstrated, the extent of academic buoyancy can lead to completely different reactions when faced with a setback. Students like Gaius, experience more negative emotions due to setbacks and bounce back less from them. When a student experiences those negative

---

[1] *O nee, nog meer zinnen! Ik dacht dat ik ervan af was. Nou ja, misschien ben ik opeens heel goed in ze vertalen.*

[2] *[Als ik een beetje gefrustreerd ben] dan neem ik minder de tijd om alles een beetje rustig door te lezen en rustig te kijken.*

emotions, they are at risk of developing a fear of failure and avoiding future challenging situations (Fried & Chapman, 2012; Meneghel et al., 2019). However, students who are academically buoyant, like Phaedra, have been found to have high confidence levels, persist through difficulties and practice composure and control (Martin & Marsh, 2008). Thus, in being able to reach full learning potential, academic buoyant students are at a distinct advantage over those who are not buoyant.

Despite a differing extent of academic buoyancy, Phaedra and Gaius did have something in common: within this study they were both identified as cognitively gifted (for more on how we operationalised giftedness, see Chapter 2). Their giftedness makes it unlikely that they are regularly confronted with being unable to successfully fulfil school tasks (Balduf, 2009). As explained in more detail in the Introduction, in the long term, having low academic buoyancy puts students at extra risk of larger problems such as (chronic) underachievement and dropping out (cf. Alexopoulou, Batsou & Drigas, 2019). By enhancing academic buoyancy in gifted students, teachers might be able to set an important step in keeping more gifted students from dropping out or underachieving. Therefore, academic buoyancy in gifted students is of particular interest.

The fact that Gaius and Phaedra were both gifted students, but reacted differently to the setback task seems to indicate that other factors than (in)experience with being challenged might moderate the extent of academic buoyancy in gifted students. In the present study, we measured the extent of academic buoyancy in terms of behaviour. In this respect, the study differs from other studies on academic buoyancy (e.g. Kim & Han, 2014; Comerford, Batteson & Tormey, 2015; Verrier, Johnson & Reidy, 2018; Jahedizadeh, Ghonsooly & Ghanizadeh, 2019). As explained in Chapter 1, we designed the Latin Buoyancy Task in such a way that we could isolate a setback and compare the pre- and post-setback outcomes. Non-buoyant participants, such as Gaius, were recognised as those who were unable to bounce back and perform the post-setback task similarly on accuracy to the pre-setback task, whereas others were able to bounce back.

In the present chapter, we take the first step in the analysis of our verbal transcripts of the Latin Buoyancy Task and, through them, academic buoyancy. During this first step, we adopted a quantitative analysis method to answer whether and in what ways the gifted participants were affected by the setback. Moreover, we also aimed at discovering to what extent that effect was moderated by independent variables, i.e. our learner variables. We chose to start with a quantitative analysis method of the coded verbal transcripts. We aimed first to identify more specifically which participants were affected by the setback.

Furthermore, we aimed at establishing which of the learner variables moderated the extent of the effect of the setback on the participants' translation accuracy and process. In Chapters 5 and 6 we further explain what caused these changes in some participants, by qualitatively analysing the verbal transcripts.

In this first analysis, we focused on comparing the outcomes from the pre-setback task with those from the post-setback task. To measure the extent of 'bouncing back', we compared the participants' translation accuracy of the pre- and post-setback task (see Chapter 2, p. 95). We hypothesized that students with a decrease in accuracy (i.e. non-buoyant on the Latin task) would have also adapted something in their translation process (c.f. Ransdell, Levy & Kellogg, 2002). Therefore, we not only compared the accuracy of the pre- and post-setback tasks, but also three other variables that were indicative of whether the participants had adapted their translation process. Thus, we compared the accuracy of the pre- and post-setback tasks and then compared the three process indicators to operationalise the translation process.

The first process indicator we included was the use of *implicit metacognition*. We included this as a process indicator, as metacognition plays an important role in Latin translation tasks (e.g. Florian, 2015; Boyd, 2018; Luger, 2018). We measured implicit metacognition on the basis of how often the participants switched between thinking activities (see Chapter 2 for how we distinguished thinking activities and Chapter 3 for how we coded them). In doing so, we were for example following Sun & Matthews (2012). Switching from one thinking activity to another implies that the participant (subconsciously) monitored that they should be doing something else (c.f. Veenman, 2015). In the qualitative analyses of Chapters 5 and 6, we delve into explicit metacognition, but as a first quantitative exploration, we thus focus on implicit metacognition.

*Perseverance* was the second of the three process indicators. This was chosen as a process indicator because a setback during a task can lessen the motivation to persevere (Weiner, 2018). Similarly to Lucas et al. (2015), we measured perseverance on the basis of the total time on task. According to to Goldhammer et al. (2014), time on task is the amount of time that is needed to fulfil the task. We chose time on task to indicate perseverance, as it says something about how long the participants persevered after the setback task and if that was different from what they would do under normal circumstances, i.e. pre-setback. We expected a general decline of time on task as the task progressed (Honeyfield, 1993), but a significant loss of time on task might indicate an effect of the setback. A study by Naumann (2019) has shown that time-on-task effects particularly present themselves when there is a difficult task and according to Lockl & Schneider (2003), changes in time-on-task reflect changes in metacognitive activity. Moreover, in the case of gifted students, perseverance might be thwarted by fear of failure (Dai, 2000). Therefore, we included perseverance as a process indicator.

We did expect that there would be a decline in perseverance in general throughout the Latin Buoyancy Task (Arnau, Brümmer, Liegel & Wascher, 2021). However, in the case of a significant decline in perseverance, this might be indicative of the participants having changed their process. Gaius's quote "When I'm frustrated by a task, I am less bothered about reading everything and taking my time to look properly" seemed to support this hypothesis.

The final process indicator was *grit*. Grit is strongly related to the personality trait *conscientiousness* (Duckworth & Quinn, 2009). According to a definition by Maddi, Matthews, Kelly, Villarreal & White (2012), grit is the courage to persevere at something when failure is possible. They reported that gritty people are steadfast in their task approach and do not quickly deviate from that approach. Many studies have corroborated the positives of grit (e.g. Duckworth et al, 2007; Robertson-Kraft & Duckworth, 2014). However, too much grit is also known to be hindering when tasks are difficult, as they persist despite little chance of success (Moutafi, Furnham & Paltiel, 2004; Lucas, et al., 2015). Seeing the untranslatable nature of the setback task, grit was included in our study.

We measured grit on the basis of the average duration of a thinking activity (see Chapter 2). Grit was therefore a combination of the other two process factors. We divided the time on task (perseverance) by the number of switches (implicit metacognition). Thus, we knew how long the participant committed to the same activity (grit) on average.

The aim of this chapter was not only to ascertain whether and in which way the participants' translation processes were affected by the setback, but also to measure to what extent that effect was influenced by individual differences. These were measured via learner variables. Therefore, we selected three learner variables that might affect the extent of academic buoyancy in students: [1] Latin translation proficiency, [2] mindset and [3] frustration tolerance. Here we will explain why these learner variables were chosen as possible moderators of the effect of the setback task on the participants' translation processes. For the exact definitions of these learner variables, see Introduction p. 8, for how we measured them see Chapter 2, p. 95. The second aim of this chapter was, thus, to investigate which learner variables contribute to the extent of academic buoyancy in the participants.

A participant's Latin translation proficiency might influence to what extent said participant is affected by the setback task. All the participants might be gifted, but that does not automatically make them all great translators. Hom &

Maxwell (1983) concluded that the amount of effort put into a task is influenced by the extent of the success or failure the participant expects based on the task difficulty. This holds especially true for participants with high Latin proficiency in translating. These participants should be more aware of task difficulty, whereas for less able students the relatively easy pre-setback task might already be perceived as difficult. Being conscientious of failing can lead to negative thoughts about oneself and distract from the task at hand (Baumeister & Tierney, 2012; Nordman & Adcock, 2022). Moreover, students can also evade failure by dysfunctional strategies when they believe a task is too difficult for them (Boekaerts & Niemivirta, 2000; Sobocinski et al., 2020). Therefore, the perceived difficulty of the setback task might affect the participants' buoyancy.

Mindset (e.g. Dweck, 2011) was the second learner variable we included as a possible moderator of academic buoyancy. For, participants with a growth mindset might persevere more and experience fewer negative emotions (King, 2017). In turn they are likely to have developed their metacognitive skills more, as to do so practice with challenging tasks is necessary (e.g. Veenman, 2008). Moreover, they are more attentive to feedback regarding learning strategies (Blackwell, Trzesniewski & Dweck, 2007). Furthermore, depending on how students view challenges, they might avoid them or engage extra in them (Efklides, Kourkoulou, Mitsiou & Ziliaskopoulou, 2006; Zimmermann & Schunk, 2011). Therefore, mindset seemed a learner characteristic that might affect the participants' extent of academic buoyancy.

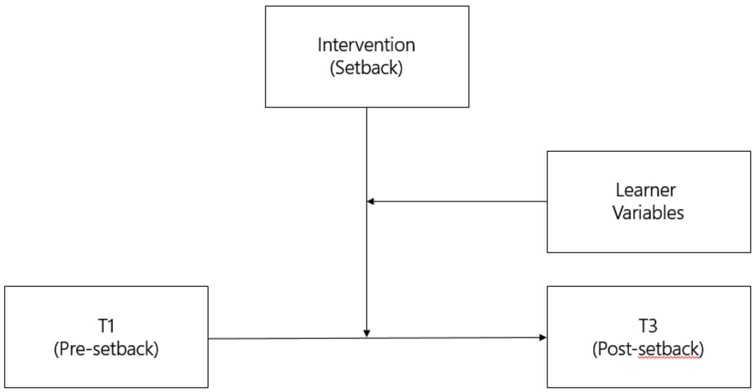
The final learner characteristic was frustration tolerance. The extent of frustration tolerance is known to influence academic achievement. Students often experience frustration during academic tasks (D'Mello, 2013). Students with a low frustration tolerance are more likely to avoid challenges (e.g. Solomon & Rothblum, 1984; Bridges & Roig, 1997; Harrington, 2005; Wright, Lam & Brown, 2009). In the case of gifted students in particular, there might be an asynchronous development: at a young age they were often able to quickly pick up new skills. Consequently, their frustration tolerance is not always adequately developed which can cause problems during puberty (Silverman, 2002; Grobman, 2006). We, therefore, also added frustration tolerance as a possible moderator of academic buoyancy in our study.

In sum, the current chapter forms a first step in the analyses of the data on academic buoyancy. Using a quantitative approach, we answer whether and in which way the low buoyant participants were affected by the setback task. Moreover, we aimed at explaining differences regarding the extent of academic buoyancy on the basis of differences in learner variables. To do so, we compared and contrasted the results from the low buoyant participants with those from the academic buoyant participants. We hypothesized that, in the case of participants who were low buoyant on Latin translation, the decline in translation accuracy would accompanied by a change to their translation process which was brought about by the setback. We expected that low buoyant on Latin participants would adapt their process post-setback by using less implicit metacognition, persevering less and/or displaying less grit. Furthermore, we expected that these alterations would interact with the learner variables. The findings from this chapter are further explored qualitatively in Chapters 5 and 6.

## 2 METHOD

For this dissertation, we measured the extent of academic buoyancy in terms of task behaviour using the Latin Buoyancy Task. We made use of a mixed-methods design to interpret the data. The model we used for this part of the study can be found in Figure 4.1: we monitored the participants before and after the intervention. We specifically compared and contrasted the accuracy and translation processes of the pre- and post-setback tasks to describe the effect of the intervention. We then investigated whether the effect of the intervention was moderated by the three learner variables.

*Figure 4.1. Model of the quantitative analysis*




### *2.1 Participants*

Besides Gaius and Phaedra, this study included 14 other gifted participants, whose parents had given active consent for participation. All 16 participants were in their third year of Latin. The mean of the participants' Latin grades was 6.8 (*sd* = 1.4), similar to the national average.

As demonstrated in Chapter 2, we had purposefully selected these participants from a larger pool of students. One of the selection criteria for participation was that they were either particularly non-academic buoyant or otherwise particularly buoyant, while performing the Anagram Task. Based on this, we expected them to also be particularly buoyant or non-buoyant on the Latin Buoyancy Task. Ten of our participants were non-buoyant participants on the Anagram Task, the other six were, thus, academically buoyant on that task. Table 4.1 presents these participants, their translation academic buoyancy, their Latin translation proficiency, their self-reported mindset preference and their self-reported frustration tolerance. For the exact scores regarding these variables, we refer to Chapter 2, p. 99.

*Table 4.1 List of participants and their learner variables*

| | Learner variables | | | |
|---|---|---|---|---|
| Pseudonym | Anagram: Academic buoyancy | Translation proficiency | Mindset preference | Frustration tolerance |
| Quintina | Low | Low | NP[1] | Average |
| Diana | Low | Average | Growth | High |
| Octavia | Low | Average | Growth | High |
| Julia | Low | Average | Growth | Low |
| Rufus | Low | Average | Fixed | High |
| Anthony | Low | High | Fixed | Average |
| Gaius | Low | High | Fixed | Average |
| Ennius | Low | High | Growth | Average |
| Flavia | High | Low | Fixed | High |
| Marcus | High | Low | NP | Average |
| Phaedra | High | Average | Growth | Low |
| Livia | High | Average | Growth | Average |
| Nona | High | Average | Fixed | High |
| Horatia | High | High | Growth | Low |

*Note.* [1] *NP stands for no preference.*

### *2.2 Materials*

For the qualitative analysis, we made use of the Latin Buoyancy Task, the Mindset and Frustration Tolerance questionnaires and the Anagram Task. We described these instruments extensively in Chapter 1. In Table 4.2 we present a summary of the instruments that were of interest for the analyses of this chapter and how they related to the learner variables. For reasons explained in Chapter 3, we made use of the codes from main coding scheme for the qualitative analysis.

*Table 4.2 Summary of materials*

| | | Buoyancy | | Learner variables | | | Process indicators | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Anagram | Latin | Proficiency | Mindset | Frustration tolerance | Implicit mc[1] | Perse-verance | Grit |
| Anagram Task | | x | | | | | | | |
| Questionnaire | | | | | x | x | | | |
| Latin Buoyancy Task | Transcripts pre-setback task | | x | x | | | x | x | x |
| | Setback task | | | | <<<Intervention>>> | | | | |
| | Transcripts post-setback task | | x | | | | x | x | x |

*Note.* [1] *Mc stands for metacognition.*

### *2.3 Procedures*

All the procedures followed for the data-collection can be found in Chapter 2.

### *2.4 Analyses*

To answer whether the learner variables accounted for the differences between the participants' academic buoyancy, we conducted a regression analysis. We did so as a regression analysis is used to verify if there is a relationship between multiple variables (Buijs, 2017). When using regression analysis, the researcher is aiming to predict the effect of the independent variable on the dependent variable (Buijs, 2017). In our case, we wanted to know whether all three or any of the learner variables (translation proficiency, mindset and frustration tolerance) moderated the effect of the setback on the students' translation accuracy and processes. We operationalised academic buoyancy via accuracy. Thus, translation accuracy was our first dependent variable and the learner variables formed our independent variables.

As we hypothesised that any effect of the setback would be related to them altering something in their translation process, we also included the three process indicators (switches for implicit metacognition; time on task for perseverance; average thinking-activity duration for grit) as dependent variables. By including three independent variables and in total four dependent variables, there were 12 possible relationships for which we created regression models.

The models we created and analysed were separate nested models. Creating nested models was appropriate for our study, as we had collected data using repeated measures (pre-setback and post-setback) using the same individual participants (Field, 2017). By using nested models, we were able to demonstrate whether the variables predicted the setback's effect on buoyancy and on the participants' translation process better individually or combined. Using a nested model analysis would also allow us to study whether there was an interaction effect between the learner variables and the effect of the setback.

The nested models we created and analysed using SPSS v.25. They existed of the base model and a further three models:

*Model 0.* For the base model, we included two elements. The first item we included was the so-called intercept. The intercept represents the mean of the dependent variable in the sample when no explanatory variables are taken into consideration. Secondly, we included a random factor, in our case: individuals. While the participants provided scores for a variable both pre- and post-setback, the scores are nested in individuals, and the within subject's variation will vary. This component corrects the error variance due to variation within participants.

*Model 1: Effect of intervention*. The first model investigated whether there was an effect of the setback on the scores on a dependent variable. In other words: did scores on, for example, the participants' translation accuracy differ before and after the setback intervention?

*Model 2: Effect of learner characteristics*. The second model answered if qualities of the subjects explain variance in the dependent variable, next to the effect of the intervention (Model 1). Or, put differently: did the participants' translation proficiency, their mindset preference, or their frustration tolerance explain variation in translation processes?

*Model 3: Moderating effect of learner characteristics*. In the third and final model, we added the interaction between the factors from Model 1 and Model 2. By doing so, we were able to answer whether learner characteristics moderated the setback's effect on the dependent variables. We shall exemplify this using mindset as the independent variable and translation accuracy as the dependent variable. The third model then answered whether the effect of the setback task on the participants' translation accuracy varied according to the participants' mindset preference. Or, in other words, was the effect of the setback larger when participants were more inclined to one of the mindset preferences?

## 3 RESULTS

The models were nested and, therefore, the model fit could be directly compared. The best-fitting model for each learner characteristic can be found in Tables 4.3. The estimates can be found in Appendix 4.1.

*Table 4.3 Model comparisons for Accuracy, Number of Switches, Time on Task, and Activity duration*

*Accuracy*

| Models | | Comparison | | | | | |
|---|---|---|---|---|---|---|---|
| | | *X^2* | *df* | Models | *X^2* | df | *p* |
| 0 | Intercept & random factor | 3941.38 | 1 | | | | |
| 1 | Plus: Intervention effect | 2827.046 | 2 | 0 vs 1 | 1114.334 | 1 | <.001 |
| Effect of Learner variable: Translation proficiency | | | | | | | |
| 2 | Plus: Lv[1] effect | 2793.976 | 3 | 1 vs 2 | 33.070 | 1 | <.001 |
| 3 | Plus: Interaction | 2129.394 | 4 | 2 vs 3 | 664.582 | 1 | <.001 |
| Effect of Learner variable: Mindset | | | | | | | |
| 2 | Plus: Lv effect | 2826.827 | 3 | 1 vs 2 | 0.22 | 1 | 0.639 |
| 3 | Plus: Interaction | 2808.368 | 4 | 2 vs 3 | 18.459 | 1 | <.001 |
| Effect of Learner variable: Frustration Tolerance | | | | | | | |
| 2 | Plus: Lv effect | 2826.827 | 3 | 1 vs 2 | 0.22 | 1 | 0.639 |
| 3 | Plus: Interaction | 2808.368 | 4 | 2 vs 3 | 0.732 | 1 | 0.392 |

*Number of switches*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | Intercept & random factor | 8405.53 | 1 | | | | |
| 1 | Plus: Intervention effect | 7457.904 | 2 | 0 vs 1 | 947.626 | 1 | <.001 |
| Effect of Learner variable: Translation proficiency | | | | | | | |
| 2 | Plus: Lv effect | 7455.783 | 3 | 1 vs 2 | 2.121 | 1 | .145 |
| 3 | Plus: Interaction | 7444.642 | 4 | 2 vs 3 | 11.141 | 1 | .001 |
| Effect of Learner variable: Mindset | | | | | | | |
| 2 | Plus: Lv effect | 7455.222 | 3 | 1 vs 2 | 2.682 | 1 | .101 |
| 3 | Plus: Interaction | 7248.835 | 4 | 2 vs 3 | 206.387 | 1 | <.001 |
| Effect of Learner variable: Frustration Tolerance | | | | | | | |
| 2 | Plus: Lv effect | 7457.011 | 3 | 1 vs 2 | .893 | 1 | .345 |
| 3 | Plus: Interaction | 7452.997 | 4 | 2 vs 3 | 4.014 | 1 | .045 |

*Time on Task*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | Intercept & random factor | 14392.91 | 1 | | | | |
| 1 | Plus: Intervention effect | 13447.67 | 2 | 0 vs 1 | 945.231 | 1 | <.001 |
| Effect of Learner variable: Translation Proficiency | | | | | | | |
| 2 | Plus: Lv effect | 13447.232 | 3 | 1 vs 2 | .442 | 1 | .506 |
| 3 | Plus: Interaction | 13437.928 | 4 | 2 vs 3 | 9.304 | 1 | .002 |
| Effect of Learner variable: Mindset | | | | | | | |
| 2 | Plus: Lv effect | 13447.67 | 3 | 1 vs 2 | .002 | 1 | .964 |
| 3 | Plus: Interaction | 13306.33 | 4 | 2 vs 3 | 141.343 | 1 | <.001 |
| Effect of Learner variable: Frustration Tolerance | | | | | | | |
| 2 | Plus: Lv[1] effect | 13446.16 | 3 | 1 vs 2 | 1.517 | 1 | .218 |
| 3 | Plus: Interaction | 13424.88 | 4 | 2 vs 3 | 21.277 | 1 | <.001 |

*Average Thinking Activity Duration*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | Intercept & random factor | 5635.356 | 1 | | | | |
| 1 | Plus: Intervention effect | 5575.255 | 2 | 0 vs 1 | 60.101 | 1 | <.001 |
| Effect of Learner variable: Translation Proficiency | | | | | | | |
| 2 | Plus: Lv effect | 5574.968 | 3 | 1 vs 2 | .287 | 1 | .592 |
| 3 | Plus: Interaction | 5573.094 | 4 | 2 vs 3 | 1.874 | 1 | .171 |
| Effect of Learner variable: Mindset | | | | | | | |
| 2 | Plus: Lv effect | 5572.472 | 3 | 1 vs 2 | 2.783 | 1 | .095 |
| 3 | Plus: Interaction | 5546.075 | 4 | 2 vs 3 | 26.397 | 1 | <.001 |
| Effect of Learner variable: Frustration Tolerance | | | | | | | |
| 2 | Plus: Lv effect | 5575.252 | 3 | 1 vs 2 | .003 | 1 | .956 |
| 3 | Plus: Interaction | 5496.682 | 4 | 2 vs 3 | 78.57 | 1 | <.001 |

*Note*. [1] LV stands for learner variable.

The run models indicated that the intervention had generally affected the participant's translation accuracy, use of implicit metacognition, perseverance and grittiness. In the case of accuracy, use of implicit metacognition and perseverance, the effect of the setback was negative. This meant that, on average, the participants' translations were less accurate after the setback compared to before the setback. Moreover, they spent less time on the task and switched less between thinking-activities. For grit, we found a positive effect: typically, the participants spent longer per thinking-activity after the setback. Thus, the setback task affected the participants' accuracy and their translation process.

For translation accuracy, proficiency did account for differences between the participants' scores. In the other cases, the learner variables had no significant effect. However, in most cases, an interaction effect was found between the learner variable and the intervention. In some cases, the interaction was positive. This meant that when the participants scored higher on the learner characteristic, the effect of the intervention was larger. This, for example, applies to accuracy and proficiency: when the participants' proficiency was higher, the effect of the setback (i.e. decline in accuracy) was larger. In the other cases, a negative effect was found: when the participants scored lower on the learner characteristic, the effect of the intervention was larger. An example of a negative interaction effect was found between accuracy and mindset: when the mindset score was low (i.e. indicating a fixed mindset preference), the effect of the setback on the participants' accuracy (decline) was large. In Table 4.4 we summarize whether the interaction effect was negative or positive for the different variables.

*Table 4.4 Summary of interaction effects.*

| | Accuracy | Task Time | # Switches | TA |
|---|---|---|---|---|
| Proficiency | Positive | Positive | Negative | Neutral[1] |
| Mindset | Negative | Negative | Negative | Positive |
| Frustration Tolerance | NS | Positive | Neutral | Positive |

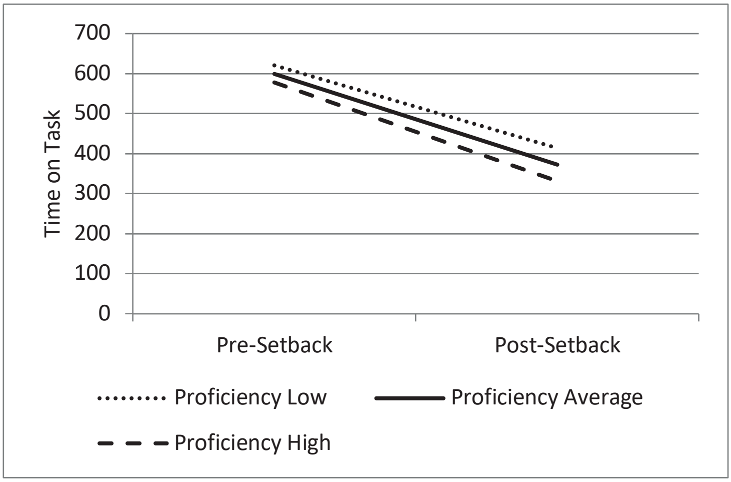
*Note. [1] Neutral indicates no significant interaction effect.*

In the following paragraphs, we present what Table 4.4 summarizes per learner variable. In the case of proficiency, the results indicated that when the participants were more proficient, the effect of the setback was greatest. For, their accuracy declined the most post-setback, their time on task declined the most (see Figure 4.2 below), also, the amount they switched between thinking activities declined the most. There was no significant interaction effect found for the average duration of the thinking activities: all participants, regardless of their translation proficiency skills, increased their average duration.

For mindset, we found an interaction effect for accuracy and all three process indicators. In the case of accuracy, time on task and the number of switches used, the interaction was negative: when the participants leaned more towards a fixed mindset preference, the effect of the intervention was larger. We, thus, found that their accuracy declined the most, they spent the least amount of time on task and switched the least between different thinking-activities after the setback. For an example of this, see Figure 4.3, which represents the interaction effect for mindset and time on task. For the thinking activity duration, the interaction effect was positive; when the participants leaned towards a fixed mindset score (i.e. a low score), the duration of the participants increased the least.

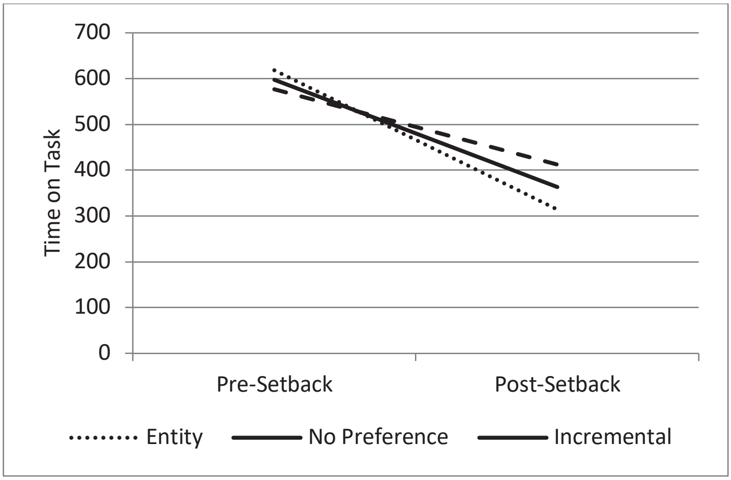
In the case of frustration tolerance, we only found interaction effects in the case of time on task and the average duration of the thinking activities. When the participants had a high frustration tolerance, both the decrease in the time spent on task (see Figure 4.4) and the increase in the duration of the thinking activities post-setback was the largest.

*Figure 4.2. Proficiency vs Time on Task.*



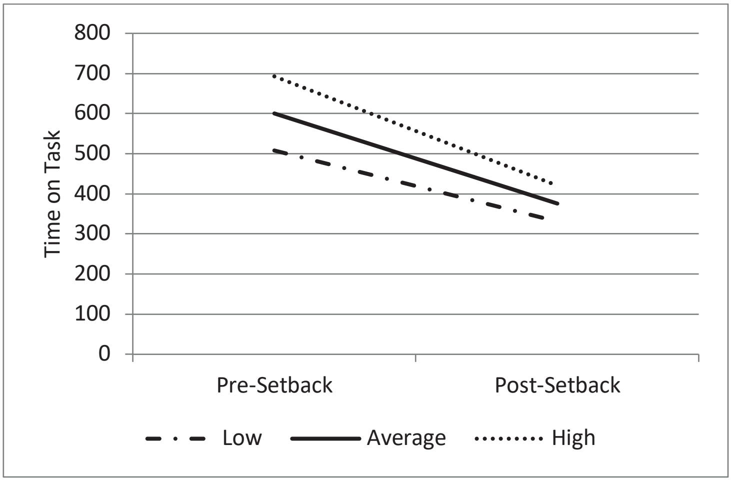

*Note*. Time on task is represented in seconds.

*Figure 4.3. Mindset vs Time on Task.*




*Note*. Time on task is represented in seconds.

*Figure 4.4. Frustration Tolerance vs Time on Task.*




*Note*. Time on task is represented in seconds.

## 4 DISCUSSION

We expected that the learner variables proficiency, mindset and frustration tolerance would moderate the effect of the setback on the participants' translation accuracy. Particularly, when a participant was a highly proficient translator, had a fixed mindset preference, or a low frustration tolerance we expected the effect of the setback to be larger. Thus, we expected the learner variables to be related to the participants' academic buoyancy on the Latin Buoyancy Task. The results showed that proficiency and mindset did indeed moderate the effect of the setback regarding the decrease in accuracy in the post-test: when the proficiency was higher or the mindset preference was less inclined towards a growth mindset, the decrease was larger. However, frustration tolerance played no role in explaining the setback's effect on the accuracy of the participants' translations.

Moreover, we hypothesised that any effect of the setback on the participants' accuracy would be paired with a change in their translation process. According to the results for translation proficiency and mindset, this was the case. Even in the case of frustration tolerance, which did not seem to moderate the effect of the setback task on the accuracy, changes regarding the participants' translation process in the post-setback task were observed. In the coming sections, we discuss the results sorted by learner variable.

### *4.1 Academic buoyancy and Latin translation proficiency*

We found a general decline in translation accuracy after the setback. The extent of this decline was found to be related to the participants' translation proficiency: when the participants had a higher translation proficiency, they bounced back less to their pre-setback translation accuracy compared to those with a lower translation proficiency. Thus, the effect of the setback was particularly large when a participant was a proficient translator under normal circumstances. Despite the trend of translation accuracy declining after the setback, in the case of the least proficient participants, their accuracy improved. The setback thus had for some a positive and for others (i.e. the non-buoyant participants) a negative effect on our participants. We supposed that the setback had led them to alter something in their translation process.

In the case of the highly proficient participants, the reason that they altered their translation process might have been that they were aware of the increased difficulty of the setback task (Hom & Maxwell, 1983; Baumeister & Tierney, 2012). The setback was thus larger for them, (c.f. Norman, 2020). A possible explanation might be that they were not satisfied with their unusually poor translation result and adopted dysfunctional strategies (Boekaerts & Niemivirta, 2000; Sobocinski et al., 2020) such as giving up on translating properly (Tulis & Fulmer, 2013). An indication that this was indeed the case, might be found when looking at the effect the setback had on the highly proficient participants' perseverance. For, after the setback, they spent significantly less time on the task, than before the setback. Where prior to the setback they had on average spent 577.88 seconds translating, post-setback the highly proficient participants only spent 331,29 seconds doing so. As with the accuracy, there was a general loss of time spent on task. However, there was also a positive interaction effect found. This meant that for the highly proficient participants, this loss was the largest. Therefore, it seems that the highly proficient participants in particular were less able to persevere after the setback, possibly indicating they gave up.

An alternative or additional explanation for the highly proficient participants' accuracy having diminished post-setback, is that they were more easily satisfied with their outcome. This could have led them to evaluate their progress less securely than before the setback. For, the use of metacognitive activities such as monitoring and evaluating is particularly triggered by difficult tasks (Veenman, 2011). Also, when confidence grows, students' monitoring and evaluation become less thorough (Molenberghs, Trautwein, Böckler, Singer & Kanske, 2016). If the participants felt the task was easy, they might have (unconsciously) been less inclined to monitor their progress.

Support for this assumption might be found in their display of implicit metacognition: the decrease in switches was the largest for the highly proficient participants. Not only did they switch less than before the setback, which is logical seeing that they spent less time on task, but, they also spent longer on each thinking activity before switching compared to before the setback. In the case of grit, there was no interaction effect, so all participants spent longer on the same thinking activity. However, in combination with the large loss of time on task, it did imply that the highly proficient participants were translating differently after the setback and that this was likely connected to a change in their use of (implicit) metacognition.

Our results not only indicated a change in the accuracy and the translation process of the highly proficient participants due to the setback, but we also found a change in that of the less proficient participants. Noteworthy was that their accuracy in some cases improved post-setback. The effect of the setback on their translation process regarding perseverance and implicit metacognition was smaller than in the case of the highly proficient translators. To some extent, this might be explained as these participants might have been less sensitive to the extreme differences in difficulty, as the pre-setback task was already quite challenging to them (Norman, 2020). But the actual improvement did imply the presence of at least some academic buoyancy in these participants: they did not give up on the task and adapted their translation process for the better. This observation led to questions such as: did these participants adapt their translation process as they went along and learned on the spot, or was the translation process of our weaker participants just more capricious?

These were not the only unanswered questions after the quantitative analysis of the data. Overall, the highly proficient translators seemed more affected by the setback. To explain this fully, the quantitative analysis was insufficient: did the highly proficient participants give up on trying their best or did they overestimate the accuracy of their translation more after the setback? Did the setback lead to the participants applying dysfunctional strategies? And, what happened during the setback to cause this effect? To further strengthen our insights into academic buoyancy on Latin translation tasks and how this relates to proficiency, a qualitative analysis of the verbal transcripts was necessary, not only of the pre- and post-setback task but also of the setback itself.

### *4.2 Academic buoyancy and Mindset preference*

The participants' mindset preference also moderated the effect that the setback had on their translation accuracy. The participants who had a fixed mindset preference were most affected by the setback: their translation accuracy showed the greatest decline after the setback. Thus, when the participants were more inclined towards a growth mindset, they were more academically buoyant on the Latin translation task. This finding agreed with the existing body of literature relating to mindset, which has demonstrated that students with a growth mindset are less affected by challenges (e.g. Hong, et al., 1999; Yeager & Dweck, 2012).

The change in the fixed mindset participants' accuracy in the post-setback task was also reflected by their translation process. In each of the three process indicators, the largest effect was found for the participants with a fixed mindset preference: these participants showed the largest decline in their time spent on the task and in their number of switches between thinking-activities during the post-setback task, compared to before. In the case of the thinking-activities themselves, the average duration of them increased the most in the fixed mindset participants.

How did their mindset preference account for the changes to their translation process after the setback? Students with a growth mindset are known to be more intrinsically motivated to complete tasks than students with a fixed mindset (e.g. Mueller & Dweck, 1998; Cury, et al., 2006). Fixed mindset students are also known to exert less effort on difficult tasks compared to students with a growth mindset in a variety of tasks (Saunders, 2013; Celis Rangel, King & Muldner, 2020). This might be what the larger decrease in time on task indicates for the participants with a fixed mindset preference: the fixed mindset participants put less effort into the task after not being rewarded for their effort, resulting in them spending less time on the remaining task.

The lower motivation levels in the participants with a fixed mindset might also be visible in the implicit metacognition results. For, being less motivated has been proven to affect metacognition, particularly lessening monitoring behaviour (Efklides, Schwartz & Brown, 2018). Furthermore, the setback might have led participants with a fixed mindset to feel negative emotions (King, 2017), which in turn are known to negatively affect metacognition (Efklides, Schwartz & Brown, 2018). Thus, the large decrease in the number of switches between thinking-activities in our participants with a fixed mindset might be indicative of them changing their translation process as a result of the setback.

Finally, the fixed mindset participants were also the participants who increased the average duration of their thinking-activities the most. This indicated that these participants had indeed changed something in their translation process besides time on task. For, they not only switched less due to them spending less time on task but they also switched relatively less.

Spending longer doing the same thing, on the one hand might be construed as positive. In that case, the participant is not giving up, but trying to push through something. However, that is to some extent unlikely in the case of participants with a fixed mindset, unless said participants were afraid to stop trying out of fear of failure. On the other hand, continuing the same thinking activity for longer might be indicative of something negative: the participants were no longer using their full problem-solving toolbox to come to a translation. Qualitative analysis of the verbal transcripts might help determine what was causing them to spend longer on the same thinking-activity.

### *4.3 Academic buoyancy and Frustration tolerance*

The effect of the setback on the participants' translation accuracy was not moderated by their frustration tolerance. The participants' accuracy generally decreased. However, we found no significant difference between participants with low or high frustration tolerance, despite the overlapping characteristics of students with a high tolerance and what is needed for academic buoyancy.

We included frustration tolerance, as we expected the untranslatable nature of the setback task to cause frustration in some participants. After all, frustration is an emotion that can be caused by failure (Berkowitz, 1989). It is particularly triggered in stressful situations (e.g. Mahon, Yarcheski, Yarcheski & Hanks, 2007; Seymour & Miller, 2017; Sorrenti et al., 2019). Possibly the participants did not experience stress and, therefore, their accuracy was not affected by the setback. The setting might for at least some part accounted for this. To maximize the quality of the think-aloud data, the setting was made to be as stress-free as possible (see Chapter 2). This might have affected the extent of the frustration felt by the participants. Another possibility is that the successful completion of the task was not an important enough aim for the participants to become frustrated (Harrington, 2011; Meindl et al, 2019).

Despite not having found a moderating effect of frustration tolerance on the participants' translation accuracy, we did find an effect on their translation process: switches decreased, implicit metacognition decreased and thinking activity duration increased. Moreover, in the case of perseverance and grit, the effect of the setback on the translation process was moderated by frustration tolerance. Ghisi, Bottesi, Re, Cerea & Mammarella, (2016) indicated that low frustration tolerance can lead to motivation loss. In addition, these students give up more easily during difficult tasks (Hoza, et al., 2001). This might be what the larger loss of time on task in our study indicates. As was the case with mindset, the increase in the average duration of the thinking activities is more difficult to interpret without a qualitative analysis of the verbal protocols.

There are, thus, indications that frustration tolerance did have some effect on how the participants approached the task after the setback. However, the participants' translation accuracy was not notably affected by a low frustration tolerance as expected. Moreover, in hindsight, there are doubts whether we were able to accurately measure behaviour related to frustration tolerance. Therefore, for the remainder of this dissertation, we do include the emotion 'frustration' as an affective response to the Latin Buoyancy Task, but we do not further include 'frustration tolerance' as a learner variable to distinguish individual diffences.

In sum, for proficiency and mindset, we found that the participants' translation accuracy was affected by the setback. This was more so the case when the participants were highly proficient or inclined to a fixed mindset preference. The results also seemed to confirm the notion that the decline in accuracy was related to an adaptation of the translation process. Besides finding these answers, new questions also arose from the data. An example of such a question was whether the highly proficient participants indeed gave up more easily or overestimated the quality of their translation accuracy post-setback. The current quantitative analysis was not sufficient to answer such questions. Therefore, a qualitative analysis of the verbal protocols was be necessary to gain a further understanding of the participants' academic buoyancy and how the setback affected them. In the next chapter, we, therefore, qualitatively analysed the verbal transcripts of the high and low proficient participants and those with an outspoken mindset preference, specifically looking at what happened during the setback.

By also qualitatively analysing the data, we could tackle one of the limitations of this chapter's study. For this study, we focused on implicit metacognition and adopted a quantitative approach to measuring metacognition. However, in the case of metacognition, it is known that qualitative analysis produces richer insights (Van der Stel & Veenman, 2008). An example of the limitedness of the current analysis is that it does not present any information on the quality of the participants' use of metacognition. We now know that the setback affected metacognitive behaviour in some manner, providing an additional reason to further explore the effect. Chapters 5 and 6 focus on the qualitative analysis. By including this further exploration, the drawback of the current chapter's quantitative approach to metacognition remains limited.

A second limitation might be found in the number of participants. 16 participants might seem on the meagre side, but as explained in Chapter 2, this number was deemed sufficient for multiple reasons, including the use of extreme sampling. An indication that we had indeed included sufficient participants, can be found in the results of the current chapter. Despite 'only' including 16 participants, we found significant effects of the setback and significant interaction effects. Thus, the effect size must be large, otherwise, the effects would not have been found to be significant with 'only' 16 participants (Sullivan & Feinn, 2012). Therefore, the small sample size is less of a limitation than might seem at first sight.

Despite this only being the first step in analysing the verbal transcripts, the current observations did already provide insights for Latin teachers into academic buoyancy during translation tasks. If their students are to learn to translate heterogeneous authentic Latin texts well, it seems that their students must be specifically taught how to cope with the changes in sentence complexity within texts. For, our results seem to imply that good translators for some reason or other are less well able to bounce back after a difficult passage. This is particularly pressing as third-year Latin students who are proficient translators, seem to be the least academically buoyant whilst translating, just as they are at the brink of the transition to translating authentic Latin texts with all their inherent changes in complexity. Latin might be a chance to let gifted students practice and eventually grow at being academically buoyant, but, to do so, there must be specific attention paid to this phenomenon by teachers during the lessons.

## 5 CONCLUSION

In the classroom context of reading and translating Latin texts, students encounter both difficult and easy sentences, in a random order (i.e. as they occur in the text). The results of this study may be discussed in this light: teachers need to anticipate the buoyancy it takes to translate a Latin text. Furthermore, Latin teachers need to bear in mind that good Latin students are not automatically buoyant students. The quotes in the introduction of this chapter from Gaius and Phaedra illustrate this.

Gaius was a participant who was non-buoyant on the Latin task: his accuracy post-setback deteriorated quite significantly. In the context of the results from the quantitative analysis his behaviour was exemplary: he had reported a fixed mindset preference on the one hand, and on the other, his proficiency was above average. Gaius' citation included at the introduction was of particular interest, due to its reference to time on task: he was fully aware that he had changed something in his translation process because he felt frustrated that he had successfully completed his second task. In contrast, Phaedra was buoyant during the Latin Buoyancy Task as her quote suggested: her accuracy did not decrease significantly post-setback. Again, this concurred with her lower proficiency and her being more inclined to a growth mindset. By examining more than only these two example quotations, further insights into academic buoyancy should be attainable.

Further qualitative examination of the verbal transcripts should provide more information on how exactly the participants were affected by the setback. This might be particularly valid regarding their translation process: in this study, for example, only implicit metacognition was considered. It would be interesting to examine which kind of thinking activities the participants were using and whether these changed after the undoable task. Knowing what they were doing differently could lead to the design of principles for strengthening their buoyancy while translating. This analysis can be found in Chapter 6.

Before moving on to that analysis, we first specifically analysed the translation process of the participants during the setback. These findings are presented in Chapter 5. In that chapter, we aimed at answering what happened to the participants and their translation process during the setback task. By doing so, we hoped to not only be able to say that participants were affected differently by the setback but also explain why that was. In both Chapters 5 and 6 we shall delve deeper into the translation processes of Gaius, Phaedra and others.

# CHAPTER 5

# PROCESSING THE SETBACK TASK: A QUALITATIVE ANALYSIS

Where Chapter 4 focused on comparing the pre- and post-setback tasks quantitatively, in the current chapter, the setback task itself is scrutinized qualitatively. We aimed at answering how translation proficiency and mindset preference account for differences between the participants' display of metacognitive and affective strategies throughout the setback task. Moreover, we aimed at confirming that proficiency and mindset were independent moderators. To do so, we created visualisations of the translation process of participants. We first compared and contrasted the so-called process bars of the high and low proficient participants. We found that the translation process of the high proficient participants altered the most throughout the setback task. The results suggested that the relationship between the effort put into the task and the reward is larger for these participants than for the low proficient participants. However, this still requires further research (see Chapter 6). We also compared and contrasted the translation process of the growth and fixed mindset participants. Here we found that the coping strategies of the fixed mindset participants were less helpful than those with a growth mindset. This chapter forms an important next step in our exploration of academic buoyancy. For, by focussing on what happened to the participants during the setback task, we also gained more understanding of what happened after the setback.

# 1 INTRODUCTION

In the previous chapter, we quantitatively analysed the think aloud data from the Latin Buoyancy Task. We found that participants were affected differently by the setback. Those who were proficient translators or held a fixed mindset preference did not return to their pre-setback translation accuracy. The data also suggested that they had altered their translation process as a consequence of the setback task. In this current chapter, we take a closer look at the setback task. We try to ascertain what led some of the participants to have significantly changed their translation process in the post-setback task compared to in the pre-setback task, ultimately leading to their translation accuracy changing as well.

Learning processes consist out of three types of strategies: cognitive, metacognitive and affective (e.g. Vermunt, 1992). Previous think aloud Latin translation studies have mainly focused on cognitive strategies when translating, often in relation to the question which strategies are employed by good translators (e.g. Florian, 2015; Boyd, 2018; Luger, 2020). However, our study is not interested in what strategies lead to the best translations, but in academic buoyancy. To overcome a setback, and thus be properly buoyant, one must first be aware (metacognitive) of experiencing (affective) a setback. Cognitive strategies might be used and affected in reaction to their task awareness and emotions (Efklides, Schwartz & Brown, 2018; Zhang & Zhang 2022), but are not key in determining how students are affected by a setback. The Latin Buoyancy Task (see Chapter 1) was created to mimic the heterogenous nature of Latin texts, including a difficult passage, in other words, a setback. In this current chapter, we, therefore, included the demonstrated metacognitive and affective strategies as the main translation process variables.

As we had found that translation proficiency and mindset preference moderated the effect of the setback task on the participants' ability to bounce back (see Chapter 4), we included these learner variables again in the current study to differentiate between participants. By doing so, we could confirm that proficiency and mindset preference function independently of each other regarding their effect on the participants' translation process. Moreover, it brought us to the following research question: how do translation proficiency and mindset preference account for differences between the participants' display of (cognitive), metacognitive and affective strategies throughout the setback task?

To answer this question, we created the contextualizing coding scheme (see Chapter 3) to map the participants' translation process of the setback task. For the current research question this coding scheme seemed the most relevant as it aims at coding for underlying strategies. In the following sections we further present the relevance and operationalisation of translation process variables 'metacognitive strategies' and 'affective strategies'. In the case of affective strategies, we explicitly pay attention to the concept of perseverance and grit. We then include the hypotheses we formulated.

### *1.1 Metacognitive strategies*

Besides the fact that to overcome a setback one must be metacognitively aware of that setback, metacognitive strategies are also relevant to this study due to the difficulty of the setback task. For, metacognition is particularly triggered when a task becomes difficult (e.g. Veenman, 2011). Moreover, the affective codes gain extra meaning when combined with the metacognitive ones. For, task difficulty perception and the ensuing (unconscious) decision to fight (i.e. persevere) or flee (i.e. avoid) is a result of metacognitive monitoring (Efklides, Papadaki, Papantoniou & Kiosseoglou, 1997; Zimmermann & Schunk, 2011; Winne, 2018; Sobocinski et al., 2020; Malmberg, Haataja & Järvelä, 2022). Thus, metacognitive strategies are of particular relevance for this chapter.

In the previous chapter, we focused only on implicit metacognition by measuring it via the number of switches between thinking activities (Sun & Matthews, 2012). However, that did not lend us the opportunity to study and review the different types of metacognitive strategies the participants employed. Using the coded verbal transcripts, that *is* what we do in the current chapter. The added value of this approach of studying metacognition has been stressed in other studies (e.g. Veenman, Elshout & Groen, 1993; Veenman, Prins & Verweij, 2003; Veenman 2011; Young & Worrell, 2018; Al Qahtani, 2020). In this chapter we, therefore, specifically focus on the metacognitive utterings themselves.

### *1.2 Affective strategies*

Affective strategies cover all strategies that are related to the regulating of emotions that occur when learning. As explained in Chapter 3, we bundled all emotions together under either the code 'negativity' or 'positivity'. Moreover, we included codes to operationalise perseverance on the one hand and grit on the other.

For, where studies including linguistic challenges for studying the educational effects of challenging students seem rare (see Introduction, p. 6 ff), difficult linguistic tasks can be found as instruments used to study perseverance or grit and their relation to challenges. The term grit demands a specific definition, especially as the term is used for different constructs in the literature. Generally, grit is connected to long-term goals (e.g. Duckworth & Gross, 2014). In that interpretation, it is a specific form of perseverance and seems pivotal for gifted students to be able to fulfil their potential (Duckworth, et al., 2007; Duckworth 2009). We, however, follow another definition of grit. In this interpretation grit is the ability to overcome the fear of failure, as found in Maddi, et al. (2012) and Lucas, et al. (2015). This interpretation of grit allowed researchers to create lab settings to study grit by prompting failure via challenging tasks, which were sometimes linguistic.

One such study was created by Lucas, et al. (2015). They measured 'grittiness' by how long the participants tried solving individual anagram tasks, including unsolvable ones. Through this, they demonstrated that fear of failure plays an important role in the student's decision to give up on the task or persist. They concluded that students with little grit gave up more quickly and very 'gritty' students can be unwilling to stop, even when solving the task at hand is hopeless. Another study related to perseverance and grit which used difficult anagrams was conducted by Gerhards and Gravert (2020). To operationalise perseverance, they measured how often their participants gave up on trying to solve an anagram. They found a significant relationship between Duckworth, et al. (2007) grit questionnaire and perseverance during the Anagram Task.

In the previous chapter, two of the variables we included to indicate an effect of the setback task on the translation process were also *perseverance* on the one hand and *grit* on the other. In that chapter we operationalised perseverance via time-on-task and grit via the average length of their bursts of thinking activities. These measurements are included again in the current chapter. Additionaly, in this chapter we also include the affective codes of the utterances as indications of perseverance and grit in our analysis. The affective codes in this coding scheme focused on underlying emotion (*negativity* or *positivity*), on motivation loss and on avoidance. Similar to Gerhards and Gravert (2020), we interpreted motivation and avoidance loss as an indication of a lack of *perseverance* and *grit* respectively.

### *1.3 Hypotheses*

In Chapter 4 we hypothesised that when the participants were more proficient, they also experienced the setback task as a larger setback due to them being more aware (i.e. a metacognitive action) of the difficulty increase (c.f. Efklides, et al., 1997; Sobocinski et al. 2020). In the current study, we hoped to ascertain if this was indeed the case. If this was the case, we would expect indications of this in the affective codes. They might refer more often to the task difficulty and/or be more negative than the low proficient participants, as task perception is known to negatively affect student's engagement in the task (Greene, Hutchison, Costa & Crompton, 2012). If our current study of the setback task indeed indicated that the high proficient participants were more aware of the difficulty, this would go some way in explaining why their accuracy suffered so significantly after the setback (see Chapter 4).

For the current study we also formulated other hypotheses. Here we first present the hypotheses related to the difference between the translation process of the high and low proficient participants. This is followed by our hypotheses regarding mindset. The hypotheses are then summarized in Table 5.1 at the end of this section.

We formulated four other hypotheses concerning proficiency. First, we expected that the proficient translators' verbal transcripts would generally demonstrate relatively more and more varied metacognitive strategies than those of the low proficient translators. For, good metacognitive skills are of particular interest for successfully performing school tasks and thus related to high proficiency (Veenman, 2008) and expert learners are known to monitor more regularly (Pressley & Afflerbach, 1995; Veenman, Kok & Blöte, 2005).

Secondly, we expected the high proficient participants' metacognitive strategies to be more affected towards the end of the setback task than those of the low proficient participants. For, their heightened awareness of the difficulty might lead to a production deficiency (Veenman, Kerseboom & Imthorn, 2000; Veenman, 2013).

Thirdly, we expected that the indications of perseverance and grit would be less present in the high proficient participants towards the end of the task compared to earlier in the task. For low proficient translators we expected less of such a change as the task progressed. This hypothesis was based on the findings in Chapter 4, where we saw that in the post-setback perseverance and grit had been significantly affected by the setback task: they displayed less of it. Moreover, we hoped to explore this further as Straka, Portešová, Halámková and Jabůrek (2021) found that in slightly younger participants that a lack of confidence led gifted participants to spend more time on the task (i.e. display grit).

Finally, we hypothesised that generally the translation process of the high proficient participants would be more structured than that of the low proficient participants. Due to their low proficiency, these participants might be seen more as novices and novices are known to work in an unstructured manner (Brown & Pressley, 1994; Pressley & Afflerbach, 1995; Veenman, Kok & Blöte, 2005).

For mindset we formulated a further four hypotheses. First, we expected the growth mindset participants to generally maintain their translation process more throughout the task than the fixed mindset participants. We based this hypothesis on our conclusions from Chapter 4. Secondly, we expected that the participants with an outspoken fixed mindset preference to generally monitor less than those with a growth mindset. For, they are known to reflect less upon their work (Nussbaum & Dweck, 2008). Thirdly, we hypothesised that where the growth mindset participants would display frequent indications of perseverance and the fixed participants would regularly demonstrate avoidance strategies. We expected this based upon the results from studies by Hong et al., (1999), Blackwell, Trzesniewski & Dweck, (2007), Duckworth, (2009) and Lucas et al. (2015), which all indicated fixed participants avoiding challenges and growth participants engaging and persevering when challenged. Finally, we expected less indication of negativity and giving up in the growth mindset participants, due to students with a growth mindset being more likely to have positive coping skills (Hong et al., 1999; Cook & Artino, 2016). These hypotheses are summarized in Table 5.1.

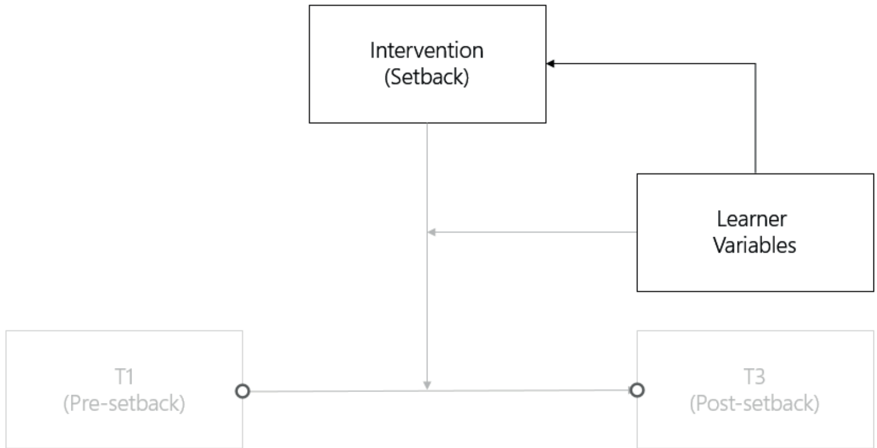
*Table 5.1 Summarized hypotheses*

| | Proficiency | Mindset |
|---|---|---|
| General | High proficient participants would display a more structured process | Growth mindset participants would maintain their process throughout the task |
| Metacognitive | High proficient participants would display:<br>1. More and more varied metacognitive strategies at first;<br>2. A larger effect on meta-cognitive strategies | Growth mindset participants would display more monitoring. |
| Affective | High proficient participants would display:<br>1. More references to difficulty and negativity;<br>2. A lessening of perseverance and grit as the task progressed. | Growth mindset participants would generally display:<br>1. Frequent displays of perseverance and grit;<br>2. Less negativity and giving up. |

# 2 METHOD

For this dissertation, we measured the extent of academic buoyancy in terms of task behaviour using the Latin Buoyancy Task. We made use of a mixed-methods design to interpret the data. The model we used for this part of the study can be found in Figure 5.1: we monitored the participants' translation process during the setback task. We specifically compared and contrasted the translation process of the most proficient translators with the least proficient translators on the one hand. On the other hand, we compared and contrasted the translation process of the participants whose mindset preference was the most outspokenly fixed to the process of those with the most explicit growth preference. We specifically focused on the participants' (lack of) displays of metacognitive and affective learning strategies.

*Figure 5.1. Model of the current qualitative analysis*




## 2.1 Participants

The 16 participants who were introduced in Chapter 4 were also used to select ten participants for the current study. We particularly focused on the participants who were either the most or least proficient translators on the one hand, and on the other hand on the participants who had the most outspoken mindset preference.

Just like we did in Chapter 4, we used the pre-setback task to ascertain the participants' translation proficiency. For this chapter we defined a participant 'high proficient' when the participant had translated the pre-setback task perfectly. When a participant had a maximum of two out of six points for the pre-setback translation, that participant was included as a 'low proficient translator'. To select the participants with the most outspoken mindsets, we looked at their mindset scores; the three participants with the highest scores were included as growth participants and the three with the lowest scores were included as fixed participants. The participants and their scores alongside other known data are presented in Table 5.2.

Using these scores, Anthony, Ennius and Gaius proved to be the most proficient participants and Flavia, Marcus and Quintina the least proficient, Anthony, Gaius and Rufus were considered as having the most outspoken fixed mindset. Horatia, Livia and Phaedra were classed as having the most prominent growth mindset preference. Anthony and Gaius were both among the most proficient participants and those having an outspoken fixed mindset preference and thus fulfilled a double role in this study. This makes the question of whether proficiency and mindset are different phenomena with different effects on the process, particularly relevant.

*Table 5.2 Participants*

| | | | Learner variables | | | |
|---|---|---|---|---|---|---|
| Pseudonym | Translation Accuracy[1] | Mindset Mean[2] | Category[3] | | Emoji 3[4] | Subjective Mark[5] |
| | | | Proficiency | Mindset | | |
| Gaius | 6 | 2.33 | High | Fixed | Irritated | 1 |
| Anthony | 6 | 2.67 | High | Fixed | Uncertain | 4.5 |
| Ennius | 6 | 4.67 | High | | Confused | 4 |
| Horatia | 5 | 5.33 | | Growth | Uncertain | 5 |
| Rufus | 4 | 2.33 | | Fixed | Uncertain | 4.5 |
| Nona | 4 | 3 | | | Uncertain | - |
| Claudia | 4 | 4 | | | Frustrated | 4.5 |
| Octavia | 4 | 4.33 | | | Interested | 3 |
| Diana | 4 | 4.67 | | | Shy | 5.5 |
| Livia | 4 | 5.33 | | Growth | Confused | 8 |
| Bella | 3 | 4.33 | | | Uncertain | 4 |
| Julia | 3 | 5 | | | Frustrated | 5 |
| Phaedra | 3 | 5.33 | | Growth | Uncertain | 4 |
| Flavia | 2 | 3 | Low | | Curious | 4 |
| Marcus | 2 | 3.33 | Low | | Frustrated | 1 |
| Quintina | 1 | 3.70 | Low | | Frustrated | 2 |

*Table notes.*
*[1]The participants' translation accuracy in the pre-setback task was scored to a maximum of six points. For more, see Chapter 1.*
*[2]These mean scores were arrived at from the participants' answers to the Dutch mindset questionnaire. For more, see Chapter 1.*
*[3]For proficiency this includes the participants who translated the pre-setback task the most and least accurately. For mindset, this includes the participants with the most outspoken preferences.*
*[4]These were the Emoji that the participants had marked prior to starting the setback task and after completing it on the Emoji Chart, see Chapter 1.*
*[5] After each task, the participants gave their translations a score out of 10. In the Netherlands a 5.5 and higher indicate a passing score.*

### *2.2 Materials*

Latin Buoyancy Task, and then specifically the setback task itself, was central to the study of the current chapter. For more on this instrument, see Chapter 1.

For the coding we made use of the second coding scheme. We developed the second coding scheme to fit the study's aim better than the first scheme (see Chapter 3). It fitted better because by using the codes that intended to convey the intent of the utterance and not just its content, we would be provided with more understanding of the participants' translation process. We also thought it would inform us more about the interaction between the cognitive, metacognitive and affective strategies.

### *2.3 Procedures*

The procedures followed for the data-collection can be found in Chapter 2. All procedures related to the compiling and coding of the verbal transcripts can be found in Chapter 3.

For this chapter perseverance or a lack of perserverences was of particular importance. To operationalise this concept, we made use of three types of data. First, we used time on task, the same measurement as in Chapter 4 . Secondly, task completion was taken as an indication of perserverence. The verbal protocols also were a data source for indications of perseverance . For example, one participant uttered 'I am through with it'[1] and finished the task. This segment was coded as 'loss of motivation' and thus a lack of perseverance. We thus used multiple methods to distinguish indications of perseverance.

Besides perseverance, grit was also of particular importance in this study. Again, we used multiple ways of establishing indications of grit or lack thereof. First, we included the average thinking-activity duration, as in Chapter 4. Finally,

---

[1] *Ik ben er klaar mee*.

the affectively coded verbalizations could be interpreted as indications of grit or a distinct lack of grit. Bella for example said 'try again'[1], showing grit by not giving up after multiple tries having tried to solve the problem at hand. In this manner we operationalised grit in the current study.

### *2.3 Analyses*

To map the process of the participants during the undoable task, we created process bars as visual schemes of their process according to the coded verbal transcripts. This visualization of the process is shown for the task as a whole and is divided into three sections: one section per sentence. This made it not only possible to compare the translation processes of different participants, but also to compare strategies within the same participant as the task progresses. The process of the first sentence indicates what the participants generally do when attempting a difficult translation task. The second and third sentences show what happens to their translation process when the difficulty is prolonged.

Using Ennius as an example, we present how we created the process bars and how we used them to analyse the participants' translation process, using all the data contained in the visualisation. The process bar visualizing Ennius' translation can be found in Figure 5.2, the corresponding colour key in Table 5.3 and in Table 5.4 the amount of words Ennius spent relatively on each type of strategy are presented.

The process bar shows that Ennius spent 1596 seconds on the task as a whole and 19 thinking activities were displayed, using a total of 437 words (see A). Of these words, 31.35% was related to cognitive strategies, 56.22% to metacognitive strategies and 22.43% to affective strategies (see B).

Each thinking activity is represented by a coloured block and classified as either a cognitive, metacognitive or affective thinking activity. In the visualisation the size of the coloured blocks does not represent the length of that strategy. The length of each thinking activity is indicated by the number of words printed in the bottom row. Figure 5.2 shows that Ennius attempted to translate all three sentences as all three sentences are accounted for in the bar. The sentences were approached in the order they were presented with no backtracking between the sentences.

[1] *Nog een keer proberen.*

*Figure 5.2. The process bar representing Ennius' translation process*

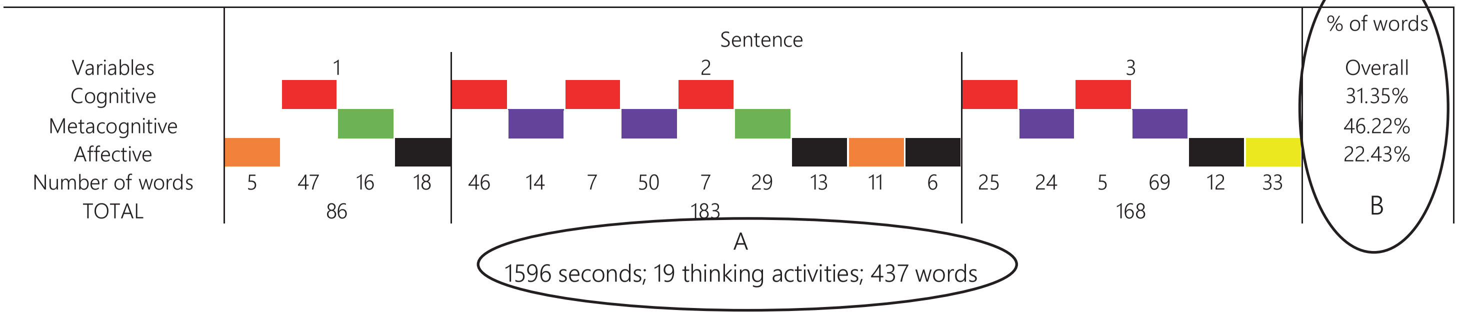



*Table 5.3 Colour key corresponding to the process bars*

| Cognitive | Metacognitive | Affective |
|---|---|---|
| Cognitive | Monitoring<br>Problem-defining<br>Stepping out | Avoidance<br>Giving up<br>Negativity<br>Positivity |

*Table 5.4 Relative words used per strategy*

| | Relative amount of words used per sentence | | |
|---|---|---|---|
| Strategy | Sentence 1 | Sentence 2 | Sentence 3 |
| Cognitive | 54.65% | 32.79% | 17.86% |
| Metacognitive | 18.60% | 50.82% | 55.36% |
| Affective | 26.74% | 16.39% | 26.79% |

By reading from left to right we can follow Ennius' translation process: Ennius started with a negative emotion towards the task, which involved translating three more sentences ('oh no not more' [1]) and then started translating with a long burst of cognitive strategy (47 words) followed by two shorter bursts of metacognition (monitoring) and problem avoidance. At this point we arrive at the vertical line in the bar, indicating that Ennius stopped translating the first sentence here and began trying to translate the second sentence.

During this second sentence, we see that Ennius increased the time spent on the sentence as a whole compared to the first, indicated by the number of words spent on the sentence. The first cognitive burst of translating lasted just as long as in the first sentence. The process bar indicates that Ennius used a new type of metacognitive thinking activity; defining a problem, saying 'I do not really know how I must turn all of this into a sentence. I am also not exactly sure what morphological ending this is [...]'[2]. From then on, Ennius shortly alternated between cognitive strategies and defining problems. Then Ennius monitored his activity and suggested a strategy he could apply if it had been a normal setting: 'Oh yes, I cannot really make much of this one either. I would now be at the point that I would go to the teacher to ask something'[3]. This was followed by a display of an avoidance strategy before an utterance with a negative meaning: 'I find this second set difficult'[4]. Ennius finally avoided the problem again saying 'I will shortly move on to the next one'[5]. This was coded as avoidance as Ennius never actually returned to try to solve the problem.

For the third sentence, the first cognitive strategy was similar in length to that of the first sentence, but later cognitive utterances were particularly short (e.g. 'you have, uhm, you have'[6]). Again, this was followed by an alternation between cognitive strategies and problem defining. By the third sentence cognitive strategy (25 words) had nearly halved compared to the first sentence (47 words). Ennius displayed another example of avoidance, this time in the form of external attribution for the struggle, 'it has been a really long time since we dealt with the it ending'[7]. Ennius then brought the task to an end by giving up: 'Yes, I am not

---

[1] *Oh jeetje, o nog meer.*

[2] *Ik weet niet zo goed hoe ik dit allemaal moet een zin van moet maken. Ik weet ook niet precies wat dit voor uitgang is het is [...].*

[3] *Oh ja, ik kan hier ook niet echt iets van maken. Ik zou nu op het punt komen dat ik naar de docent zou gaan om iets te vragen.*

[4] *Deze set 2 vind ik wel moeilijk.*

[5] *Ik ga even naar de volgende.*

[6] *Jij hebt, uhm, jij hebt.*

[7] *Het is heel lang geleden dat we de it vorm deden.*

happy handing it in this way, because I know that I can do better if I prepare for it. But I have not prepared, so I am leaving it as such.'[1]. This was coded as 'giving up' and not 'explicitly stepping out' because this segment was followed by an utterance that indicated external attribution, thus betraying an underlying emotion driving this decision to stop.

When comparing the translation process of the third sentence to that of the second and third, we can discern a change: Ennius switched more between thinking activities after the first sentence, with the peak being found in the second sentence. As the task progressed further, the bursts of cognitive strategies became shorter. Ennius displayed two different metacognitive strategies. Problem-defining became more and more prominent, as attested by the number of words devoted to this strategy. The increase in metacognitive strategy seemed to imply that the participant became more reflective as the task progressed. In the case of this participant, we can thus conclude that the prolonged exposure of task difficulty affected the participant's approach to the task with their relying more on metacognitive strategy, particularly regarding problem defining.

Ennius did show indications of perseverance, particularly during the second sentence, where a prolonged amount of time was spent trying to translate it. An indication of grit was his completing the task and that he does so without sentence three receiving markedly less attention than sentence 2 (see the number of words for each sentence). The number of words used, the number of switches and the duration of the thinking activities is neither particularly short nor long compared to the other participants. However, there were also examples of avoidance and negativity throughout the task. All this combined seems to indicate that Ennius behaved relatively gritty.

## 3 RESULTS

Figure 5.3 includes the percentage of words used per strategy and per sentence. The process bars of the high proficient and low proficient participants can be found in Figures 5.4 and 5.5 respectively, whereas the process bars of the growth mindset participants are presented in Figure 5.6 and those of the fixed mindset participants in Figure 5.7. The visualisations of the remaining 6 participants can be found in Appendix 5.1. In the following paragraphs, we share the results in more detail. We present the results regarding metacognition first and then for

[1] *Ja ik vind het heel stom om het zo in te leveren, want ik weet dat ik het beter kan als ik voorbereid. Maar ik heb niet voorbereid dus ik laat het zo.*

affection. Each time, we split the results into two parts: one for proficiency and one for mindset.

*Figure 5.4. The relative number of words used per sentence per participant*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| High Proficiency | | | | | | | | | |
| | Anthony | | | Ennius | | | Gaius | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Cognitive | 81.05 | 35.79 | 75.0 | 54.65 | 32.79 | 17.86 | 55.32 | 90.10 | 88.46 |
| Metacognitive | 10.53 | 56.32 | 0.00 | 18.60 | 50.82 | 55.36 | 42.55 | 6.93 | 0.00 |
| Affective | 8.42 | 7.89 | 25.0 | 26.74 | 16.39 | 26.79 | 2.13 | 2.97 | 11.54 |
| Low Proficiency | | | | | | | | | |
| | Flavia | | | Marcus | | | Quintina | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Cognitive | 70.90 | 77.66 | 85.46 | 80.17 | 67.72 | 69.23 | 88.98 | 78.98 | 75.0 |
| Metacognitive | 29.10 | 10.99 | 11.45 | 19.83 | 32.28 | 0.00 | 7.99 | 9.16 | 13.89 |
| Affective | 0.00 | 11.36 | 3.08 | 0.00 | 0.00 | 30.77 | 3.03 | 11.86 | 11.11 |
| Growth Mindset | | | | | | | | | |
| | Horatia | | | Livia | | | Phaedra | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Cognitive | 100.0 | 96.76 | 100.0 | 51.79 | X | X | 79.56 | 82.81 | 100.0 |
| Metacognitive | 0.00 | 3.24 | 0.00 | 42.55 | X | X | 20.44 | 17.19 | 0.0 |
| Affective | 0.00 | 0.00 | 0.00 | 2.13 | X | X | 0.0 | 0.0 | 0.0 |
| Fixed Mindset | | | | | | | | | |
| | Anthony | | | Gaius | | | Rufus | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Cognitive | 81.05 | 35.79 | 75.0 | 55.32 | 90.10 | 88.46 | 69.14 | 28.8 | 100.0 |
| Metacognitive | 10.53 | 56.32 | 0.0 | 42.55 | 6.93 | 11.54 | 30.86 | 68.0 | 0.0 |
| Affective | 8.42 | 7.89 | 25.0 | 2.13 | 2.97 | 0.0 | 0.0 | 3.2 | 0.0 |

*Figure 5.5. Process bars of the high proficient participants*

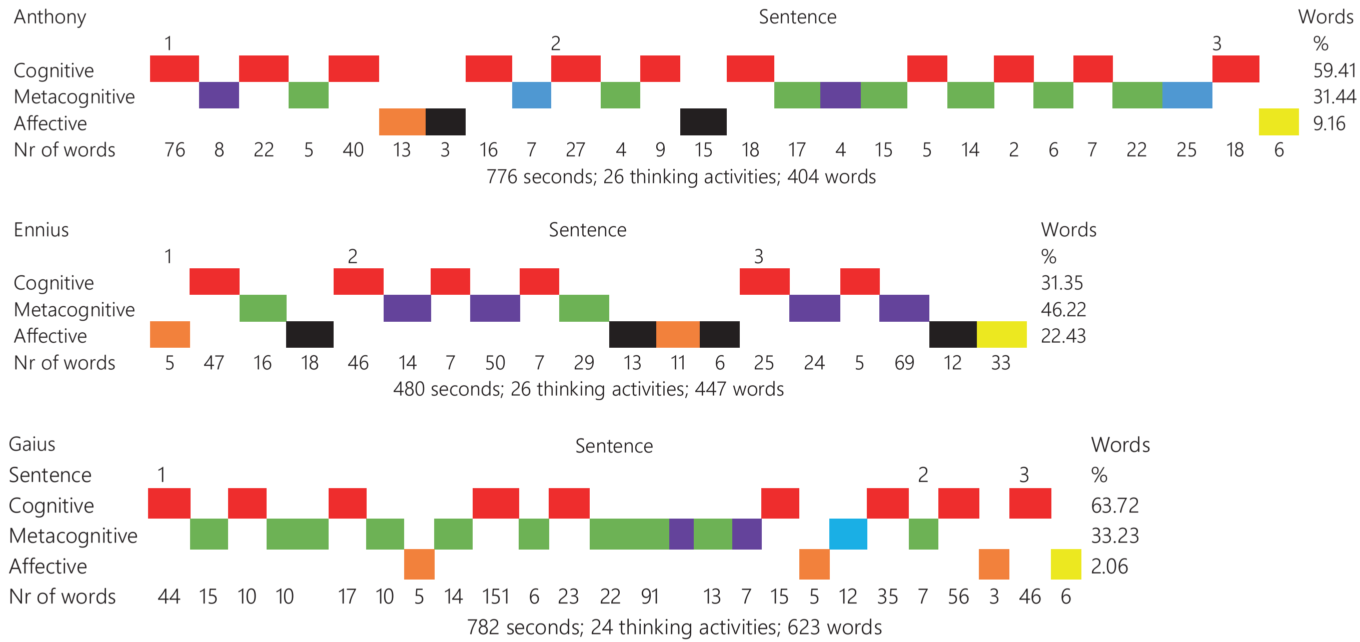


*Figure 5.6. Process bars of the low proficient participants*

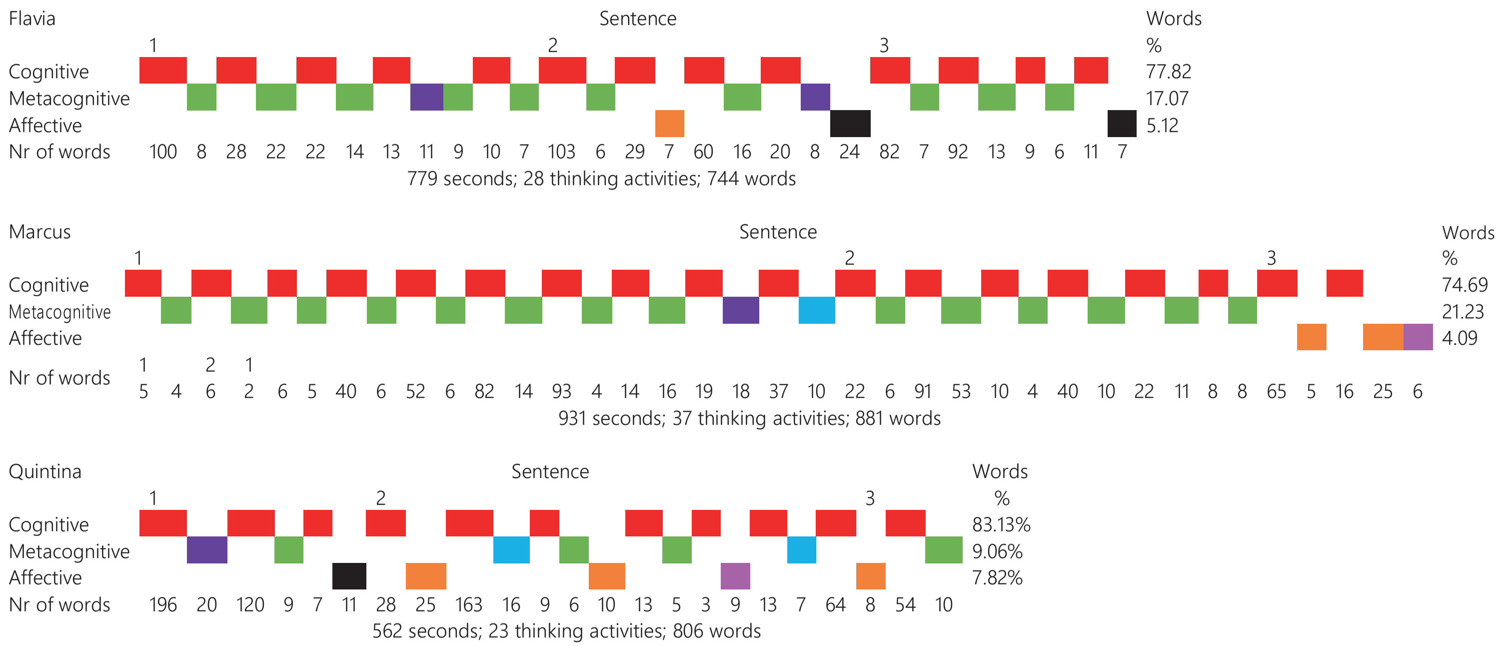


*Figure 5.7. Process bars of the growth mindset participants*

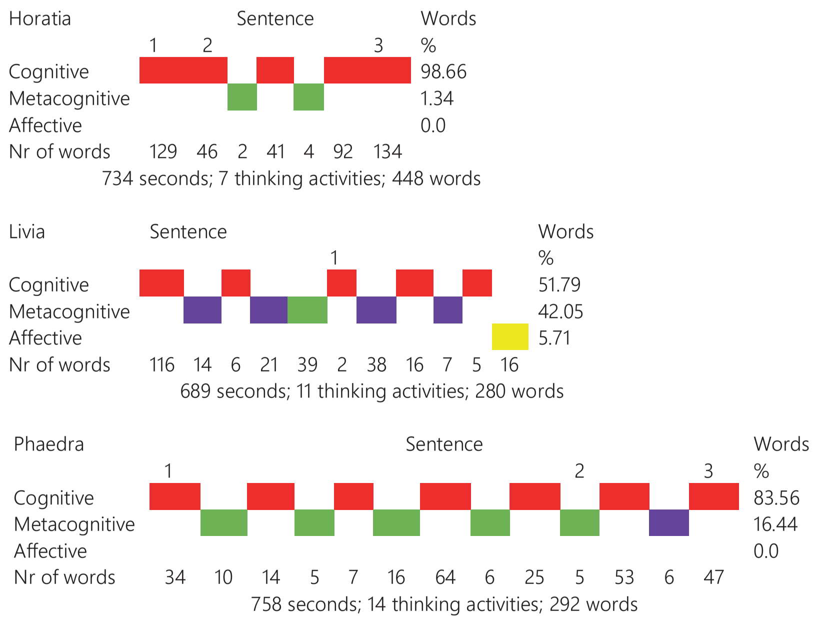
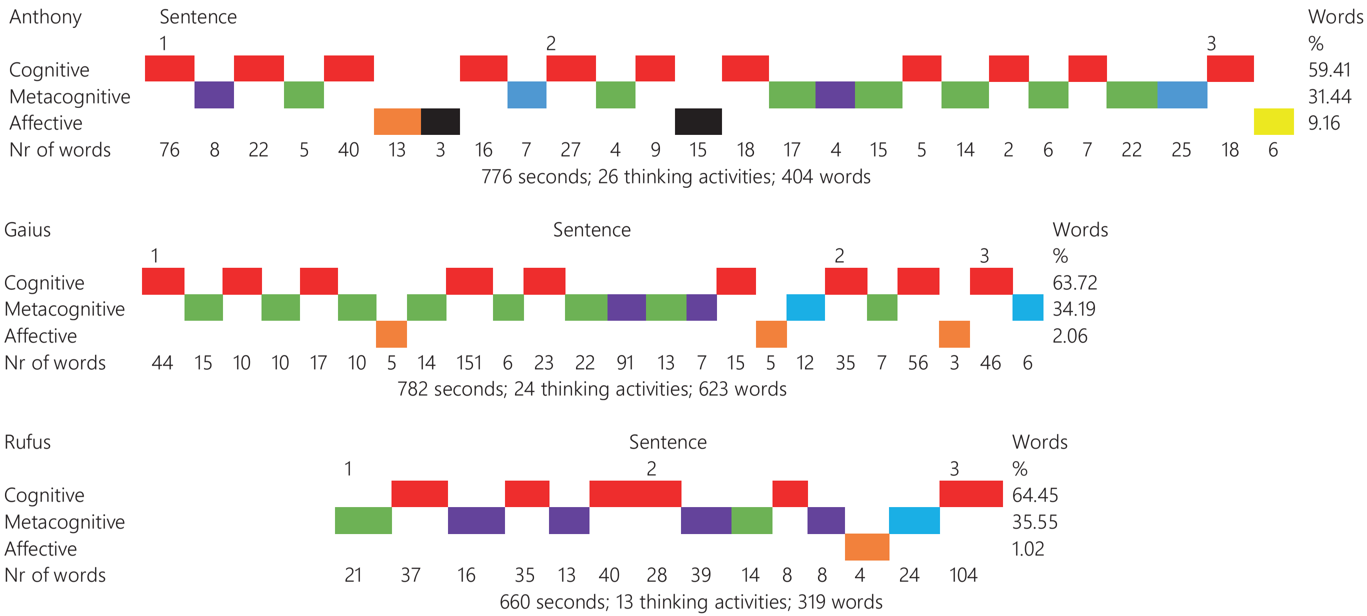


*Figure 5.8. Process bars of the fixed mindset participants*



### *3.1 Metacognitive strategies*

In the following sections we present the results related to the participants use of metacognitive strategies. For each sub-category of participants, we start with findings related to the overall relative display of metacognitive strategies. We then move on to a short summary of the process bar of each participant of one of the subcategories (e.g. high proficient), whilst comparing and contrasting their individual processes. This is repeated for the second subcategory (e.g. low proficient). Then we compare and contrast the processes of both categories of participants, including three elements: [1] relative display of metacognition in the first sentence, [2] variation in metacognition in the first sentence and [3] changes that occurred as the task progressed.

#### 3.1.1 Proficiency

Overall, the high proficient participants displayed metacognitive activity more frequently than the low proficient participants. In the case of the low proficient participants, only 9.06-21.23% of all their words were coded metacognitively, compared to 31.44-46.22% of the words of the high proficient translators.

For the high proficient participants, particularly Anthony's and Ennius' display of metacognitive strategies was comparable to each other. These participants both started in sentence 1 with relatively few metacognitive strategies (10.15% and 18.60% respectively). As the task progressed into the second sentence they not only increased their metacognitive activity regarding the absolute amount of words (Anthony: 20 >107 Ennius: 16 > 93), but also relatively. In the second sentence the metacognitive strategies accounted for more than half of their translation processes.

Gaius' translation process was slightly different. Gaius displayed relatively more metacognitive strategies during the first sentence (42.55%). However, Gaius' translation process also changed as the task progressed into the second sentence: during this sentence he made much less use of metacognitive strategies (6.93%). As the task progressed longer, Ennius' use of metacognitive strategies remained comparable to the second sentence. However, Anthony's process changed again: he no longer made explicit use of metacognitive strategies in the third sentence. Gaius also displayed no metacognitive activity in the third sentence, but this was more similar to the behaviour in the second sentence.

In the case of the low proficient participants, there was less uniformity in their translation process. Flavia's translation process started with a relatively high display of metacognitive strategies in the first sentence (29.10%), but as the task progressed into the second and third sentence this lessened to around only 11%. Marcus spent a notable long amount of time on the task as a whole (931 seconds and 881 words). Over halve of his words (479) were spent on the first sentence of which just under 20% were displays of metacognitive strategies. In the second sentence he demonstrated relatively more use of metacognitive strategies (32.28%), but then showed no metacognitive strategies in the third sentence. Quintina's process was particularly even throughout all three sentences. She did show a slight increase in metacognitive strategies as the task progressed (7.99% > 9.16% > 13.89%). The low proficient translation processes were thus not very comparable to each other.

Regarding the variety of the metacognitive activities, there was little to distinguish between the high and low proficient participants: all participants demonstrated at least two different types of metacognitive activities and as the task progressed, metacognitive activities that had not been demonstrated in the first sentence rarely occurred. For both groups, monitoring was the most seen metacognitive activity. Thus, there was no distinction regarding the variety between the high and low proficient participants.

In conclusion, the low proficient participants displayed metacognitive activities less frequently than the high proficient participants. Also, the relative changes to their translation processes were less rigorous than those seen in the those of the high proficient participants. For metacognitive variety no difference between the high and low proficient participants were observed.

### 3.1.2 Mindset

On average, the fixed mindset participants used more words related to metacognition than the growth mindset participants (38.65% compared to 19.94%). Growth mindset participant Horatia was conspicuous in her overall use of metacognition, for she displayed nearly no metacognitive activity throughout the task.

In the case of the growth mindset participants, their translation processes regarding metacognitive strategies differed greatly. Horatia, displayed no metacognitive activity in the first sentence, increased this very slightly as the task progressed to the second sentence, and then displayed none again during the third sentence. Livia only completed the first sentence. However, she spent less than a minute less on this sentence than Horatia spent on all three and more than 40% of her words were related to metacognitive strategies. Phaedra completed all three sentences. She displayed a similar amount of metacognitive activity in the first two sentences (20.44% and 17.19%), but none in the third sentence.

For the fixed mindset participants, the processes of Anthony and Rufus were relatively similar, in that both greatly increased their demonstration of metacognitive activity during the second sentence, but then demonstrated none in the third sentence. As presented in the proficiency results section, Gaius started with displaying a lot of metacognitive in the first sentence, but in the second and third sentence the metacognitive activity was significantly less.

Regarding the variety of the displayed metacognitive activities, only the fixed mindset participants showed ‘stepping out’, and moved on to the next sentence. The growth mindset participants tended to favour one type of metacognitive activity: Livia favoured problem defining; Horatia and Phaedra favoured monitoring. In the case of the fixed mindset participants, there was less obvious favouring of a specific metacognitive activity. Their displays were more diverse.

In summary, in contradiction to what we expected, the fixed mindset participants showed relatively more metacognitive activity than the growth mindset participants and did not show a particularly small variety of different activities. Moreover, the results regarding mindset indicated that the fixed mindset participants changed their translation process more regarding metacognitive activities than the growth mindset participants.

### *3.2 Affective strategies*

In the following sections we present the results related to the participants use of affective strategies, particularly focussing on perseverance and grit. We follow the same presentation order as for the metacognitive strategies. This time we also include time on task and the average number of words per thinking activity, as indications of perseverance and grit.

#### 3.2.1 Proficiency

Overall, with the exception of Ennius, the high and low proficient participants displayed relatively few affective strategies. 22.43% of Ennius’ words were coded affectively, whereas for the other participants less than 10% of their words were affectively coded. On average, the high proficient participants spent shorter ($M$ = 679.33 seconds, $SD$ = 172.65) on the setback task than the low proficient participants ($M$ = 754.00, $SD$ = 180.80). The average word count of the low proficient participants’ thinking activities was also higher ($M$ = 28.48, $SD$ = 5.86) than that of the high proficient participants ($M$ = 19.56, $SD$ = 5.60). Interestingly, the average length per thinking activity in seconds did not differ much between the low (*M* = 25.72 seconds, *SD* = 1.84) and high (*M* = 26.96 seconds, *SD* = 7.49). The high proficient participants thus used less words per thinking activity, this was particularly the case for Anthony.

When looking at the translation processes, we found that around 8% of Anthony's words included affective strategies in the first sentence. As the task progressed into the third sentence, a quarter of high proficient Anthony's words were coded affectively as he lost motivation. This was a large relative increase. Ennius is particularly interesting as he was the only participant to start his translation process with an affective code (negativity). This seemed to set the tone for the rest of his translation process. Ennius' affective strategy display remained high in the third sentence, after finishing the second sentence with a burst of avoidance and negativity. He kept displaying a lot of affective strategies and, similarly to Anthony, Ennius finished the task due to a loss of motivation. In contrast, Gaius demonstrated no affective strategies in the third sentence. In the first two sentences he did display short moments of negativity. In none of the high proficient participants did we find examples of positivity.

In the case of low proficient Flavia, the first example of an affective strategy was found in the second sentence (negativity). She ended both the second and third sentence whilst demonstrating an avoidance strategy. Marcus, despite his exceptionally long attempt at the setback task, did not display any affective strategies until his final few words. He displayed two bursts of negativity and then he ended with positivity. Quintina, who started demonstrating affective strategies at the end of the first sentence and from then on did so a little more frequently, was the only other participant to include positivity as an affective strategy. She also demonstrated avoidance strategies and negativity.

Comparing the low and high proficient participants, we find that positivity only occurred in the case of low proficiency. Negativity occurred in both low and high proficient participants, but it occurred relatively much earlier in the high proficient (on average starting at the 87$^{th}$ word instead of at the 523$^{rd}$ word). Even if we exclude the extremes of Marcus and Ennius, the difference is still large. In high proficient Anthony and Ennius we saw more avoidance than in the low proficient participants. Loss of motivation was only observed in the high proficient participants. So, the specific affective strategy display differed between the two groups.

#### 3.2.2 Mindset

Overall, the fixed mindset participants displayed marginally more affective strategies than the growth mindset participants. Regarding the setback task as a whole, the participants with a fixed mindset switched more frequently between thinking activities throughout the task than the growth participants. In addition, there was little to no difference between the participants regarding the time they spent on the task, and, therefore, we can say that the fixed mindset participants (*M* = 21 thinking activities; *M* = 22.51 words per activity) showed more but shorter thinking activities than the growth participants (*M* = 10.67 thinking activities; *M* = 31.88 words per activity).

Regarding the participants with a growth mindset, both Horatia and Phaedra demonstrated no explicit affective strategies. Livia spent nearly eleven and a half minutes only on the first sentence, which was exceptionally long. She eventually lost her motivation and despite finding it difficult, as attested by her broken voice and teary eyes, asked to move on to the next set. This led to 5.71% of her words being coded affectively, despite no other affective bursts occurring.

Avoidance and negativity occurred from the first sentence in Anthony. In the third sentence he lost motivation, and this accounted for 25% of his words related to that sentence. In comparison, no words accounted for affective strategies in the third sentence occurred in Gaius' third sentence. Short bursts of negativity did occur in the first two sentences. In Rufus' translation of the setback task we found only a negative affective code. Combining this with the above, more affect, including perseverance and grit, was seen in the fixed mindset participants.

Besides determining how the (cognitive) metacognitive and affective strategies were affected during the setback task and relating this to differences between mindset and proficiency, we also aimed to ascertain that mindset and proficiency affect the process independently. This indeed seems the case, as other differences were observed between the proficiency groups on the one hand and the mindset groups on the other.

## 4 DISCUSSION

Before discussing the results, we must scrutinize the fact the that Anthony and Gaius were included in both the high proficient and fixed mindset groups. Particularly as our participants groups were relatively small, this deserved our attention whether this had effect on the results. The fact that the growth mindset group contained only participants that were not included in the proficient groups meant that we were comparing Anthony's and Gaius' translation process to different participants and thus to different behaviour. As the participants differed we were looking at a different relationship. Moreover, this is strengthened by the fact that the participants who completed the high proficient and fixed groups were different. Therefore, it is still possible to distinguish 'fixed mindset' driven strategies from 'high proficiency' driven strategies in Anthony and Gaius. However, the comparison would have been sharper if there had been no overlap in participants. By including more participants to choose from than our original 16 participants, this might be intercepted in possible further or replicating research.

In the following sections we delve deeper into the verbal protocols to shed further light on the findings. To either support our interpretations of the process bars as presented in the Results section or to test whether they were merited, we include the emoji (see Table 5.2), the subjective marks and the quotations from the verbal transcripts, when necessary in this section. By doing so, we anticipated the further analysis presented in Chapter 6. In contrast to the presentation of the Results section, we do not discuss the findings per main strategy type (metacognitive or affective), but first by proficiency and then by mindset. We do so because we believe including the interaction between metacognitive and affective strategies strengthens our understanding and interpretation of the individual strategies.

### *4.1 Proficiency*

We hypothesised that the high proficient participants would display more and more varied metacognition compared to the low proficient participants at first and that their metacognitive activity would more affected towards the end of the setback task. We also expected that the high proficient participants would demonstrate more references to difficulty and negativity. Finally, we hypothesised that they would show less signs of persevering and grit as the task progressed. We found the following results:

1. The metacognitive activity in the first sentence was not particularly more for the high proficient participants, but when including the second sentence it was.
2. There was no difference between the variedness of the metacognitive activities of the high and low proficient participants.
3. As the task progressed, the metacognitive activity generally changed more in the high proficient participants than in the low proficient participants.
4. Regarding affective displays, the high proficient participants did not display more negativity, but when they demonstrated negativity, they did so much sooner than the low proficient participants.
5. Regarding grit, avoidance was seen most in two of the three high proficient participants. Motivation loss (i.e. a lack of perseverance) was observed in the

high proficient participants and not at all in the low proficient participants. There thus seemed less signs of persevering and grit in the high proficient participants.

Generally, we expected the high proficient participants to approach the setback task with a more structured translation process. The lack of uniformity among the low proficient participants might be in conjunction with this. Novices are known to solve problems haphazardly and without much connection between their strategies (Brown & Pressley, 1994; Pressley & Afflerbach, 1995; Veenman, Kok & Blöte, 2005).

When looking at the process bars, it seems at first sight that both the low proficient and high proficient participants mainly approached the task by alternating cognitive and varying forms of metacognitive activity. However, there was a qualitative difference: there was less connection between the cognitive and metacognitive strategies of the low proficient participants than the high proficient. For example, high proficient Anthony did not start translating straight away, but wrote down all the case endings first. He would refer to this when morphologically analysing a noun. On the other hand, low proficient Marcus monitored at one point that his translation 'does not sound Latiny'[1]. Marcus then endeavoured to conjugate, but was unable to so properly due to a lack of knowledge.

Despite the low proficient participants having displayed useful metacognitive strategies and correct observations, they did miss specific cognitive knowledge relating to Latin. As a result of this they were less able to act relevantly according to their metacognitive observations (c.f. Veenman, Elshout & Meijer, 1997; Veenman, Kok & Blöte, 2005; Cromley & Kunze, 2020). Moreover, in the low proficient participants we also found examples of a complete lack of coherence between their monitoring and their follow-up strategy that is not obviously caused by a lack of Latin knowledge. For example, Quintina said at one point when trying to determine the meaning of *cano:* 'I think dog, but it seems unlikely in this sentence. I will just go with dog'[2]. Within a sentence she went from correctly noticing that the meaning dog made little sense in the context to directly thereafter deciding to go with the meaning 'dog'. In conclusion, it seems that the difference between the low and high participants was not the metacognitive variety nor the absolute or relative amount of metacognitive strategies they demonstrated, but the quality of the connection between the cognitive and metacognitive strategies.

[1] *Dit klinkt niet 'Latijnerig'.*

[2] *Ik denk hond, maar het lijkt onwaarschijnlijk in deze zin. Ik ga gewoon hond doen.*

Where both the high and low proficient participant displayed a similar variety in different metacognitive activities, a difference did occur when further splitting the metacognitive activities: problem defining was more prominently present in the process bars of the high proficient participants. Not only was problem defining more prominent in the process bars, but also the problem analysis itself tended to be more detailed, as the following examples demonstrate:

> "The infinitive, how are we going to use that? *Delphinus* is the subject, but I do not know how I should connect it to the verb, as that is a second person and you can't connect it to you *Delphinus* or something" [1]
> - Gaius (high proficient)

> "Because promising is in the whole verb form, do not know how, yeah, need to use it."[2]
> - Quintina (low proficient)

This seems to suggest that the high proficient participants were more aware of problems and created a clearer picture of what the problem was before attempting to solve it. Again, this seems to point towards the high proficient participants having a more structured approach: a clear problem definition is an important metacognitive orientating step that helps students decide upon which strategy will best help them solve the problem (Brown & Pressley, 1994; Pressley & Afflerbach, 1995; Veenman, Kok & Blöte, 2005).

Based on the frequent metacognitive behaviour seen in the low proficient participants, it does not seem that the lower proficient participants suffered more from an availability deficiency of metacognitive strategies (Veenman, Kerseboom & Imthorn, 2000) than the high proficient participants. It also goes too far to say that the metacognitive skills displayed by the low proficient participants were poorly developed. Veenman (2015) presented that in the Netherlands 40% of the gifted students have poorly developed metacognitive skills. However, based on our results, we also cannot go as far as to say that all the participants had exceptionally well-developed metacognitive skills, which international studies have implied (e.g. Carr & Taasoobshirazi, 2008; Greene, Moos, Azevedo & Winters, 2008). More research in the Netherlands specifically addressing the metacognitive development of gifted students might cast more light on these findings, particularly in relation to their task proficiency.

In Chapter 4 we found that there was a significant decrease in the high proficient participants' translation accuracy after the setback compared to before it. We also observed that this was the result of a change to their translation

[1] *De infinitivus, hoe gaan we die gebruiken? Delphinus is een onderwerp maar ik weet niet hoe ik het moet koppelen aan de persoonsvorm. Want dat is tweede persoonsvorm en dat kun je niet koppelen aan jij delphinus of zoiets.*
[2] *Omdat beloven in het hele werkwoord staat weet ik niet hoe ik dat ja moet gebruiken.*

process in that they switched a lot less between thinking activities and displayed less indications of implicit metacognition. We hypothesised that this change might have been a continuation to what happened to their translation process during the setback. This indeed seemed the case: we found a larger change to the high proficient participants' demonstration of metacognition. The question however still remains why the high proficient participants were so affected that they changed their metacognitive process.

To answer this question, we suggest that one must incorporate the findings regarding affective strategies. Here we will anticipate this suggestion, before doing so on a larger scale in Chapter 6. Of the high proficient participants, Anthony's and Ennius' translation processes were similar in that as the task progressed into the second sentence they both seemed to increase the effort they put in the task. We interpreted effort on the basis of their increased number of words and thinking activities.

This 'effort' is reminiscent of Straka, et al.'s (2021) conclusion that gifted students increase the time they spend on the task when their confidence waives. Therefore, their increase in metacognitive activity might be a demonstration of them finding the task difficult. Ennius explicitly refers to finding the setback task difficult: ''I find this second set difficult'[1]. This interpretation might be confirmed by the scores Anthony's and Ennius' gave their own translations: for the setback task they gave themselves lower scores than for the pre- and post-setback tasks. Anthony and Ennius both gave their translations a failing score (45% and 50% respectively), where for the other two tasks they both expected passing scores. Their emoji also indicated them experiencing difficulty: uncertain (Anthony) and confused (Ennius). Combining this, with their negativity coming to light much earlier than it does in the low proficient participants, it does indeed seem that Anthony and Ennius were particularly aware of the difficulty of the task.

The initial demonstration of metacognitive strategies by Gaius seemed different than that of the other two high proficient participants. For, he mainly demonstrated them in the first sentence and spent a prolonged amount of time on this sentence. However, that prolonged time might be interpreted in the same manner as Anthony's and Ennius' second sentence: as effort exerted on account of being aware of the difficulty. Ennius rated his translation as exceptionally bad, giving it a minimum score '1' and his emoji was 'irritated'. The difference was that he was aware of the difficulty earlier than Anthony and Ennius and thus put in extra effort sooner.

---

[1] *Deze set 2 vind ik wel moeilijk.*

Thus, all three high proficient participants demonstrated putting extra effort in to translating at least one of the sentences. The extra effort that the high proficient participants put into translating might be interpreted as a demonstration of perseverance and grit, particularly as they had spent extra time on the task (Lucas, et al., 2015). However, looking at the task as a whole, and including what we know of Chapter 4, we do not think their demonstrated effort was an indication of preserving of grit. For, this extra effort was only temporary, as demonstrated by the fact that much less metacognitive activity occurred in the third sentence. Moreover, all three high proficient participants ended the task by giving up. It might be they lost their motivation due to a lack of reward for their effort (Schunk, Meece & Pintrich, 2013), where the low proficient participants placed less value on effort and ensuing result. Feeling deflated due to their effort not being rewarded with a satisfactory translation might then continue to affect their translation process in the post-setback task.

The lack of bouncing back in the high proficient participants merits further exploration, particularly of the post-setback task. Do they recover as that task progresses or not at all? It might also be interesting to research what an attainable task difficulty increase means for task process of highly proficient students, opposed to our unattainable setback task. Also, the differences between how the high proficient and low proficient exert effort whilst translating are of further interest. Furthermore, as we only focused on the setback task itself, we did not include academic buoyancy at translating as a factor in the study. This might provide further insights.

For example, an unexplored difference between Gaius on the one hand and Anthony and Ennius on the other, was that when including data from the whole Latin Buoyancy Task and not just that from the setback task, Gaius was relatively more buoyant than Anthony and Gaius. This seems to indicate that there is a difference in the effect of the setback between high proficient translators who are academically buoyant and those who are not. This difference is further explored in Chapter 6. In Chapter 6 we include the retrospective interviews, besides the verbal transcripts of the whole task of four of these participants to further explore the current findings.

### *4.2 Mindset*

With regards to mindset we hypothesised that the growth mindset participants would display relatively more monitoring activities than the fixed mindset participants. We also expected that generally they would alter their translation process less as the task progressed than the fixed mindset participants. We expected the participants who had a growth mindset preference to frequently demonstrate signs of perseverance and grit. We also expected that less negativity and giving up would occur in the participants with a growth mindset compared to the fixed mindset participants. We found the following:

1. The growth mindset participants did not display relatively more monitoring activities than the fixed mindset participants.
2. The growth mindset participants' translation process altered less as the task progressed compared to that of the fixed mindset participants.
3. On average, the growth mindset participants spent longer on the setback task.
4. Avoidance and negativity only occurred in the fixed mindset participants. Motivation loss was seen in a growth mindset participant, but only after an exceptional amount of time being spent on trying to translate the sentence.

First, to forestall any problems that might arise in interpreting the results from the fixed and growth mindset groups, we must discuss Livia's translation process further, as it was so singular. Growth mindset Livia lost motivation to continue translating the setback task after having only attempted one sentence. However, despite her giving up, we interpret Livia's strategy use as extremely gritty. Lucas, et al. (2015) concluded that being gritty led participants to be more likely to fail a larger task due to giving too much focus to one item. They also found that gritty participants put in more effort whilst losing compared to non-gritty participants. In other words, being too gritty can lead to one putting a disproportionate effort in for the achieved result. This is what we find in Livia's translation process: she spent as long trying to translate one sentence, as most participants spent on all three. By spending so long on one item, the end result was that the whole setback task would have suffered if the remaining sentences had been attainable. Thus, Livia was gritty to a fault.

In contrast to our hypothesis, the growth mindset participants did not display relatively more monitoring activities than the fixed mindset participants. In hindsight, reflecting less upon their work, as fixed mindset students were found to do by Nussbaum & Dweck (2008), does not mean that fixed mindset students do not notice problems with their work. For, to be affected by problems as fixed mindset students are known to be (e.g. Hong et al., 1999; Blackwell, Trzesniewski & Dweck, 2007; Duckworth, 2009; Lucas, et al., 2015), they must actually be aware of a possible problem. Therefore, they must be monitoring, as our results suggest.

Fixed mindset students do not differ from growth students in the action of monitoring itself, but in the affective reaction which might be related to their monitoring. For, the growth mindset participants' process bars did not contain affective strategies, except for Livia's giving up. The fixed mindset participants did, however few times, demonstrate affective strategies. When they did it was always either to avoid a problem, a loss of motivation or driven by a negative emotion. Therefore, their coping strategies when encountering problems seemed less helpful than those of the growth mindset participants (Hong et al., 1999; Cook & Artino, 2016).

The growth mindset participants do not seem to let affection drive their translation process or at least, they do so less than the fixed mindset participants. It seems that the 'distraction' of affection and not displaying helpful coping strategies to deal with said distraction, is wat led the fixed mindset participants to inadvertently change their translation process. For, the results of both the current and the previous chapters indicate that fixed mindset participants inadvertently change their process more than participants with a growth mindset.

## 5 CONCLUSION

In this chapter we aimed at ascertaining what happened during the setback task, in the hope that this might explain further why the high proficient and fixed mindset participants were not seen to bounce back after the setback task. We found indications that the role of affect seem particularly important in explaining differences between the groups of participants. We also determined that proficiency and mindset are independent moderators of the translation process when participants were translating the setback task, as comparing and contrasting the results from each individual group led to different observations.

Overall, the mindset results need little further exploration to determine how mindset affected the extent of academic buoyancy in our participants. For, their lack of helpful coping strategies and being more affectively triggered during a challenge makes it more difficult for fixed mindset participants to bounce back from the challenge. However, in the case of proficiency we have formulated new hypotheses and questions regarding their ability to deal with and bounce back from setbacks, based on what occurred during the setback task. Therefore, in Chapter 6, we shall further compare and contrast the translation process of high and low proficient participants. To do so, we look at the translation process of the whole Latin Buoyancy Task and include the retrospective thinking aloud as well.

# CHAPTER 6

# PROFILING TRANSLATION PROCESSES: FOUR CASE STUDIES

In Chapter 4 we used a quantitative method and found that when our participants were more inclined towards a fixed mindset or were more proficient translator, they were the least academically buoyant participants. In other words, their translation accuracy declined the most after the setback task. In Chapter 5 we zoomed in specifically on the setback task. We qualitatively analysed the verbal transcripts of the participants who were the most and least proficient, and the participants with the most outspoken mindset preferences. In the current chapter we zoomed in further: the focus was on four participants, each fitting a different profile, relating to their Latin proficiency (high/low) and their academic buoyancy (high/low). We aimed to contrast the strategies of participants who turned out to bounce back after the setback and with the strategeies of those who did not. Thus, we used the outcomes of the Latin Buoyancy Task, in our selection of participants for this deepened analysis. We followed these participants qualitatively throughout their progression on the Latin Buoyancy Task as a whole. Besides data from the concurrent thinking aloud, we also included data from the retrospective thinking aloud and the emoji chart as part of the analysis. Our objective was to pinpoint behaviour specific to each of our four profiles. We thus aimed to further deepened our understanding of how the participants' Latin proficiency influenced the effect of the setback task. We found that there were cognitive, metacognitive and affective effects of the setback task and that these effects depended on the Latin translation proficiency of the participants

## 1 INTRODUCTION

Thus far, we have quantitatively compared the effect of the setback on the quality of the participants' translations and their translation processes (Chapter 4). In doing so, we found that the translation accuracy decreased more after the setback when the participants' translation proficiency was higher. Subsequently, in Chapter 5, we qualitatively focused on the translation process of the participants during the setback task. In the previous chapters, we included both translation proficiency and mindset preference as moderators of the effect of the setback on their translation accuracy and processes. In this chapter, we focus on translation proficiency only. We delve into the verbal transcripts further qualitatively and build forth on what we observed the participants doing during the setback task.

In Chapter 5, we suggested that to better understand the translation process and the effect of the setback task, the relationship between the different types of strategies must be further explored. For example, both the high and low proficient participants demonstrated frequent metacognitive strategies. The data-analysis from Chapter 5 suggested that the quality of the interaction between the cognitive and metacognitive strategies was higher in the high proficient participants than in the low proficient participants. There were also indications that the high proficient participants were affected because their increased effort was not rewarded.

However, in the previous chapter, we did not yet include academic buoyancy on the Latin Buoyancy Task as a factor in the analysis. Chapter 5 focused on the participants' translation process during the setback of the Latin Buoyancy Task, and not on the translation process throughout the task as a whole. We did not yet, therefore, compare their demonstrated strategies to what they did when the task was easier, nor could we discern whether and how the setback task continued to affect their translation processes after the setback. The participants indeed experienced a setback, but how did this influence their ensuing performance?

In the current chapter, we aim to discern unique strategies as seen in the translation processes throughout the whole Latin Buoyancy Task of participants that were either high or low proficient and either high or low buoyant. We created four strategy profiles, one for each combination: [1] high proficient and low buoyant; [2] high proficient and high buoyant; [3] low proficient and low buoyant and [4] low proficient and high buoyant. We selected participants from these profiles. The translation process of each participant was compared and contrasted to those of the other participants. The strategy profiles included cognitive, metacognitive and affective strategies and the interaction between these strategies. We thus created a quadrant of strategy profiles for the strategies displayed by the participants whilst translating the three parts of the Latin Buoyancy Task.

In the current study, we focus on the difference between high and low proficient translators. This means that this study aligns most with existing studies related to Latin learning and instruction, compared to our other studies.. To our knowledge, previous studies from this field have not included the heterogeneity of the source text's complexity as a possible factor in the students' translation process (see also Chapter 3). This is all the more notable as the Latin source texts that students have to translate are so-called frustration level texts (Boyd, 2018): students must translate texts that, besides coming from a world that is unfamiliar to them, contain yet unknown vocabulary and morpho-syntactical elements. Translating and understanding Latin is thus a challenge even for the highest potentials, it is inherent to the source texts that they will frequently encounter setbacks.

Despite the importance of being able to overcome setbacks whilst translating Latin, we only found one concrete conclusion within the Latin learning and instruction literature related to difficult elements within the text. According to Florian (2015) strong translators temporarily skip difficult elements and come back to them later. We also found a study that observed two poor-performing students displaying avoidance strategies (Newland, 2016). Newland attributed their poor translations to the students' fear of failure, their lack of motivation and their suffering from learned helplessness and low self-efficacy. This is reminiscent of the behaviour of students with low academic buoyancy in that they avoid challenges due to a fear of failure (Martin & Marsh, 2008; Fried & Chapman, 2012; Meneghel, et al., 2019). This might imply that the students from that study were experiencing a setback.

Despite a further lack of references to strategies which deal with setbacks and difficulties whilst translating, much can be learnt from existing studies regarding differences between the translation processes of low and high proficient translators. For, Latin translation by students has received some attention from researchers since the 1920s. Using various methods, researchers have gained more insights into the translation process of students, often focusing on what differentiates strong from weak translators.

For Latin, think-aloud studies have concluded that proficient translators make more use of top-down strategies than weak translators and thus focus less on bottom-up strategies (Karten, 2015; Florian, 2015; Boyd, 2018). Top-down and bottom-up are both strategies translators can employ whilst translating. Top-down and bottom-up strategies are cognitive approaches to text comprehension. When bottom-up strategies are preferred, the student comes to a translation by sequentially focusing on individual words. In the case of a top-down approach, the student relies on using their background knowledge outside of the source text to come to an understanding. Karten (2015) also concluded that when low proficient translators focused on bottom-up cognitive strategies their confidence grew. However, bottom-up strategies are generally indicative of weaker translators.

Eye-tracking research confirmed that strong translators rely more upon top-down strategies than on morphological knowledge (Luger, 2018). This study also discussed that the most proficient translators were better able to switch between different cognitive strategies. Such switches can be considered implicit metacognitive skills (Sun & Matthews, 2012). The use of cognitive and metacognitive strategies seems to differentiate high and low proficient translators. The participants for these studies were generally slightly older than our participants, which might be of relevance.

Based on the knowledge from existing translation studies and our observations from Chapters 4 and 5, we formed the following hypotheses regarding what behaviour we could expect to include in the strategy profiles. These hypotheses are summarized in Table 6.1.

1. *Cognitively,* we expected our low-proficient participants to apply more bottom-up strategies throughout the whole task in comparison to high proficient participants (c.f. Florian, 2015; Boyd, 2018; Luger, 2018). For high buoyancy participants we expected the overall task process, including the use of cognitive strategies, to remain more the same than in the case of low buoyant participants (Martin & Marsh, 2008).
2. *Metacognitively,* we expected the high proficient participants to display a more useful interaction between their demonstrated cognitive strategies and what they monitored than the low proficient participants. Moreover, we expected that low buoyant, yet high proficient participants would display less useful interaction between metacognitive and cognitive strategies. For, we knew that the high proficient participants were more aware of the increased difficulty of the setback task (see conclusions Chapters 4 and 5). Thus, they were more likely to show less coherence between consecutively displayed stratetgies from the undoable task onwards and bounce back less from the setback task (c.f. Martin & Marsh, 2008; Chapter 4).

*Table 6.1 Summary of hypotheses*

| | | Academic Buoyancy on a linguistic task | |
|---|---|---|---|
| | | High | Low |
| Latin translation proficiency | High | - Makes use of top-down strategies (cognitive)<br>- Notices increased difficulty of the setback task (metacognitive);<br>- Monitoring and evaluation (metacognitive) lead to effective cognitive interventions, throughout the whole task;<br>- Is able to adjust their process and gain new insights and methods (metacognitive) as they progress through the tasks, also when experiencing difficulties (affective);<br>-Is aware of problems (metacognitive), but does not become frustrated (affective);<br>- Does not avoid problems (affective), but consciously decides to move on, (metacognitive) without further consequences (affective). | - Makes use of top-down strategies (cognitive)<br>- Notices increased difficulty of the setback task (metacognitive);<br>- Monitoring and evaluation (metacognitive) lead to effective cognitive interventions in the pre-setback task;<br>- Is unable to adjust their process and gain new insights and methods (metacognitive) during and after difficulties (affective);<br>- Experiences negative emotions during the setback task (affective);<br>- Displays avoidance strategies when encountering difficulties and experiences prolonged effects (affective). |
| | Low | - Makes use of bottom-up strategies (cognitive);<br>- Monitoring and evaluation (metacognitive) generally do not lead to effective cognitive interventions;<br>- Is able to adjust their process and gain new insights and methods as they progress through the tasks, (metacognitive also when experiencing difficulties (affective);<br>- Frequently encounters problems, but does not become frustrated (affective);<br>- Does not avoid problems (affective), but decides to move on (metacognitive) without further consequences for their mood (affective). | - Makes use of bottom-up strategies (cognitive);<br>- Monitoring and evaluation (metacognitive) generally do not lead to effective cognitive interventions;<br>- Is unable to adjust their process and gain new insights and methods (metacognitive) during and after difficulties (affective);<br>- Experiences negative emotions when encountering problems throughout the task (affective);<br>- Displays avoidance strategies when encountering difficulties throughout the task (affective). |

3. *Affectively,* differences were mainly expected depending on the participants' buoyancy. When a participant was low buoyant, we expected to find negative emotions when encountering difficulties (Martin & Marsh, 2008). They were thus more likely to become frustrated by the setback or problems

they came across when translating. These participants were also more likely to show avoidance strategies (Fried & Chapman, 2012; Meneghel, et al., 2019). This would be particularly the case when the proficiency was high, as the results suggested in Chapters 4 and 5.

To some extent, these hypotheses are evocative of Vermunt's (1996) characteristics of students' styles of learning (undirected, reproduction directed, meaning directed and application directed). However, translation is not an activity aimed at learning, so the terms are not one-on-one transferable. Still, they might provide some support in our analysis.

## 2 METHOD

### *2.1 Participants*

From the 16 think-aloud participants, four were purposively selected, based on information gathered on them and their academic buoyancy on the Latin Buoyancy Task as discussed in Chapter 3. Each participant was chosen to represent one of the four profiles presented in Table 6.1 above: [1] A high proficient, high buoyant student, [2] A high proficient, low buoyant student, [3] a low proficient, high buoyant student and [4] a low proficient, low buoyant student. Academic buoyancy was determined by how they performed on the Anagram Task and the Latin Buoyancy Task (accuracy), whereas Latin proficiency was determined by how they performed on the pre-setback task of the Latin Buoyancy Task, (see Chapter 1 for these tasks and Chapter 2 for the how we implemented these instruments).

When selecting the participants to represent the low proficient and low buoyant profile, we encountered a problem. In Chapter 1 we found that the Anagram Task for the greater part was able to distinguish students who were non-academically buoyant, with few mishits. However, of the students who were low proficient Latin translators and who proved non-buoyant on the Anagram Task, none seemed non-buoyant on the Latin Buoyancy Task. There was, therefore, no student who was a clear candidate for representing low proficient and low buoyant behaviour.

Quintina, scored much lower after the difficult anagrams than before, indicating that she did not bounce back after being confronted with a challenge. Quintina's translation of the pre-setback task was also extremely low, as it only warranted 1 point. Her post-setback task increased to three points. However, when a participant scored particularly low in the pre-setback task, there was a lot of room left for improvement, even when affected by the undoable sentences. Despite her unexpected improvement on the Latin Buoyancy Task, she still only scored 50% of the available points on the third set. This insufficient mark still indicated her being low proficient at translating Latin texts. Of all the students who continuously scored low on proficiency, this was the student with the largest decrease in score on the Anagram Task. Therefore, Quintina was selected as the low proficient, low buoyant student.

In addition to this problem, selecting a participant who was both high buoyant and high proficient also posed a problem at first sight: of the participants who scored 80 to 100% on the first Latin task, none scored equally or more after the setback task. However, Horatia only scored two points less, whereas, on average, the participants scored 2.19 (*SD* = 1.94) less points. Horatia thus fell into the normal range of decline. Her loss of accuracy might, therefore, be more perceived as a result of the task lasting longer than of the setback. Hence, she displayed a certain extent of academic buoyancy on the Latin Buoyancy Task as well as on the Anagram Task. Thus, Horatia became the representative for the high buoyant and high proficient profile.

In both the cases of Horatia and Quintina the problems in selection were based on their extreme pre-setback scores becoming less extreme post-setback. This might be an indication of regression towards the mean at the subject level. Regression towards the mean entails that in the case of repeated measures extreme scores are likely to be less extreme (and thus nearer the mean) in the next measurement (Barnett, van der Pols & Dobsen, 2004). Particularly in the case of Quintina, further investigation would be warranted in the analysis of the data to determine if she was indeed a candidate for representing low buoyancy. As we address in the Discussion, Quintina could indeed be classed as low buoyant despite the improved accuracy.

Finding participants to represent the profiles high proficient but low buoyant and low proficient but high buoyant posed no problems: multiple students qualified as possible representatives. The participants with the largest decrease or increase in scores on the Anagram Task and Latin Buoyancy Task were chosen as our case studies for the current chapter. This led to Anthony and Marcus being added to Horatia and Quintina as our four case studies to complete the quadrant. Prior to the study as a whole, their parents had been informed of the study and they supplied active consent for participation.

The key student features of each participants are shown in Table 6.2. Anthony was 14 years old. His Latin proficiency was high; however, his academic buoyancy was low on both the Anagram Task and Latin Buoyancy Task. Horatia was a 15 year old and also had a high proficiency, but her buoyancy was high. Quintina and Marcus were low proficient students of 14 years old. For the purposes of this study, Marcus was regarded as academically buoyant, whereas Quintina was not.

*Table 6.2 Participant information per student*

| | Student 1 | Student 8 | Student 11 | Student 15 |
|---|---|---|---|---|
| Pseudonym | Anthony | Horatia | Marcus | Quintina |
| Proficiency | High | High | Low | Low |
| Buoyancy | Low | High | High | Low |
| Anagram Task | -5 | +2 | +4 | -4 |
| Latin Task | -5 | -2 | +1 | +2 |

### *2.2 Materials*

We included both the verbal transcripts as they had been coded using the main coding scheme and the contextualized coding scheme for the analysis of the current chapter. By using data generated by both coding schemes, the analysis was the most exhaustive it could be.

### *2.3 Procedures*

The general procedures for the data collection are found in Chapter 2 and the coding procedures are found in Chapter 3. For this study, the translation process of the four participants was qualitatively analysed in detail for all three subtasks of the Latin Buoyancy Task. We analysed which cognitive, metacognitive and affective strategies the participants displayed. Moreover, based on what we had seen in the study focused on the setback task (Chapter 4), we also looked at whether the activities displayed after monitoring or problem defining were helpful. Besides the codes from the concurrent and retrospective think-aloud transcripts, we also included three other sources of data to form the profiles of the different participants:

1. The time spent on the task (see Chapters 2 and 4)
2. Perceived success (see Chapter 2)
3. Accuracy (see Chapters 2 and 4)
4. Emoji (see Chapter 1, Part IV).

Table 6.3 presents these data for the different participants. Scores that did not fall within a certain range (i.e., are more than a standard deviation less or higher than average), are marked with an asterisk. The average amount of time spent on task 1 was 577.63 seconds (*SD* = 215.48), that of task 2 was 651.38 seconds (*SD* = 133,78) and that of task 3 373.75 (*SD* = 124.33).

*Table 6.3 Overview of non-code variables*

| Task | Proficient<br>Buoyant | High<br>Low<br>Anthony | High<br>High<br>Horatia | Low<br>High<br>Marcus | Low<br>Low<br>Quintina |
|---|---|---|---|---|---|
| | Anagram Task | -5 | +2 | +4 | -4 |
| | Mindset | Fixed | Growth | No preference | No Preference |
| | Emoji 1 | Happy | Content | Confused | Sleepy |
| 1 | Time on task | 779 seconds | 517 seconds | 669 seconds | 345 seconds* |
| | Perceived success | 50% | 78% | 40% | 50% |
| | Accuracy | 100%* | 83.33%* | 33.33%* | 33.33%* |
| | Emoji 2 | Unsure | Unsure | Curious | Unsure |
| 2 | Time on task | 776 seconds | 734 seconds | 931 seconds* | 562 seconds |
| | Perceived success | 45% | 80% | 30% | 40% |
| | Emoji 3 | Frustrated | Confused | Interested | Unsure |
| 3 | Time on task | 450 seconds | 260 seconds | 211 seconds* | 499 seconds* |
| | Perceived success | 65% | 85% | 50% | 50% |
| | Accuracy | 16,66%* | 50% | 50% | 50% |
| | Emoji 4 | Hopeful | Confident | Happy | Curious |

*Note. * indicates that, based on the 16 participants, the score was either significantly high or low*

## 3 RESULTS

In the current section, we first verbally present a summary of the participants translation process throughout the whole Latin Buoyancy Task. These results are then summarized in Table 6.4. A full narrative account of the four participants' translation process can be found in Appendices 6.1-4, which include the concurrent and retrospective thinking aloud. These narrative accounts are supported with illustrating quotations from the verbal transcripts.

### *3.1 Anthony*

Anthony's process changed after the untranslatable task: absolutely and relatively he displayed less cognitive and even less metacognitive activity. The lack of checking and rechecking in the third task, meant he progressed through the task much quicker. Anthony felt confident when he was working quickly, but felt less confident when he was working analytically, despite the better results of that method. Noteworthy is his inability to determine correctly how well he had performed the task: in the first task Anthony underestimated his translation quite badly, where in the third task he overestimated it. During the setback task

Anthony frequently seemed frustrated, which he explicitly mentioned in the retrospective interview. He explained the frustration stemmed from the discrepancy between wanting to do well, but thinking he was not doing so and feeling he as spending too long on the task.

### *3.2 Horatia*

Overall, Horatia relied heavily on her cognitive knowledge of Latin and on little else. As the task progressed, Horatia remained analytical in her overall approach, but on a word level she became less precise and for one sentence in the post-setback task even let go of her generally morpho-syntactical driven approach. This short slip in her otherwise more steadfast translation process seemed to explain her lessened accuracy. Based on her verbalisations, she monitored little, and reflectively relied on if she was certain about the morpho-syntactical elements. After the first two tasks she reported feeling confused or uncertain, however she was satisfied after the post-setback task and she felt that she had translated the third set the best.

### *3.3 Marcus*

Generally, Marcus' translation process seemed to consist of combining logic and worldly knowledge with trial-and-error strategies. His trial-and-error method not only led to differences in his overall translation process, but also within the tasks themselves. However, trial and error strategies were observed throughout the whole Latin Buoyancy Task, with the setback task seeming to affect him little in this regard. He persisted in trying to solve the problems he encountered. However, his lack of Latin knowledge frequently hindered him from solving the problems correctly, paying more attention to grammatical features as the task progressed. Despite his self-knowledge that his translations were not correct and the change in translation process, his accuracy did not decrease, and his mood stayed positive.

### *3.4 Quintina*

Quintina's translation process heavily depended on knowing the word meanings during the pre-setback task. She seemed to create a 'bridge language' between the Latin text and the eventual translation by 'making' the Latin words Dutch and then using those Dutch words created a logical sounding sentence as translation (c.f. Bartelds, 2021). She monitored her process, but this remained superficial, and she often lacked the knowledge to adequately solve problems encountered.

She sometimes chose an option even though she knew it was wrong. However, as the tasks progressed, she adapted her process and seemed to be developing a method and trying different things.

Quintina's retrospective thinking aloud was particularly interesting. She said that she had felt disheartened whilst translating. She was afraid of her results, because she found the tasks difficult. She later mentioned that she tends to be less motivated when things were not going well and can give up on tasks. Quintina said that the fact she was not alone whilst performing the task affected her and that she thought she would have given up on the task, particularly if she had had an answer key. It seemed that under normal circumstances she would indeed have been less academically buoyant in the aspect of perseverence than the Latin Buoyancy Task results suggested.

*Table 6.4 Summary of results*

| | | Academic Buoyancy on a linguistic task: High | Academic Buoyancy on a linguistic task: Low |
|---|---|---|---|
| Latin translation proficiency | High | - Cognitively used same analytical method during setback task, became less analytical after setback.<br>- Slightly increased explicit metacognitive activity, followed by useful cognitive action during undoable task, reverted back to less explicit metacognitive strategies after the setback task;<br>- Affectively reported remaining uncertain during the setback task, despite expecting an 80% score. Reported feeling content after the third set. | - Cognitively used same analytical method during undoable task before giving up. He became less analytical after untranslatable set.<br>- As untranslatable set progressed, showed less explicit metacognitive activity, followed by some useful cognitive action. After untranslatable set, he monitored and evaluated much less;<br>- Affectively he showed and reported feeling frustrated. He gave up on the untranslatable set. Ended the third task feeling content. |
| | Low | - Cognitively, focused on logic and trial and error. Became more focused on grammatical elements as task progressed. Was unable to aptly use said grammatical elements and sometimes used incongruent knowledge to try to solve the problem.<br>- Metacognitively maintained explicitly monitoring his process throughout the setback task. After that task, he monitored, but was more negative about his abilities and outcomes. Ended on a positive note.<br>- Affectively Marcus reported being interested and intrigued by his translation throughout all three sets. | - Cognitively, became more focused on grammatical elements. Was unable to aptly use said grammatical elements and sometimes used incongruent knowledge to try to solve the problem.<br>- Explicitly displayed new metacognitive activity such as evaluation during the setback task. Continued using these ‘new’ strategies post-setback, but to a lesser degree.<br>- Reported finding it frustrating that she did not know how to translate properly. Became uncertain during the setback task, but reverted back to being interested after the third set.<br>- Indicated perseverance due to the research setting. |

# 4 DISCUSSION

An important limitation of this study is that we only included four participants to create the quadrant. However, the data collected for each participant was particularly rich, and the results seem promising. Further research including more participants fitting in the quadrant would still be advisable to strengthen and sharpen the findings. In the following paragraphs, we first discuss the results of each individual participants. Then, we group the students, first according to proficiency and then according to buoyancy.

## *4.1 Participants*

### 4.1.1 Anthony

All in all, high proficient and low buoyant Anthony had a particularly well working strategy to start with, where he combined different types of knowledge and monitored his process. He alternated cognitive and metacognitive activities, affective activities were not present. Based on his perceived accuracy, he noticed the increased difficulty of the setback task, which were the lowest for that task.

Notable was that in the post-setback task nearly no monitoring was observed. Despite him having the knowledge and ability to use this knowledge, as we had observed in the pre-setback task, his decreased use of monitoring strategies meant he scored poorly. This change in behaviour seemed fitting with low academic buoyancy. His actual low accuracy would be probably be a surprise and disappointment to him as he felt he had done well and had done his best, (c.f. Van Dijk, Van der Pligt & Zeelenberg, 1999; Pintrich, 2000; Elliot & McGregor, 2001; Byrka, Cantarero, Dolinski & van Tilburg, 2021). The disappointment might be greater as Anthony wanted to do well and became frustrated and restless when he felt it was not going well during the second set. This seemed to correspond with his lack of academic buoyancy: the change in difficulty led to negativity. Notable was that, despite his lack of academic buoyancy, he did keep trying to solve the untranslatable task, for, Lucas, et al. (2015) found that the gritty students where the ones that continued difficult tasks to a fault.

This initial perseverance might be explained by the Anthony's mindset preference and his wish to perform well. Anthony reported having an outspoken fixed mindset. Giving up on the task might have caused him to lose face, particularly as he wanted to do well. Moreover, he was focused on task completion not on task process, fitting a fixed mindset (e.g. Yeager & Dweck, 2012). Task process and the method of how he came to a translation were of minor importance to him, as supported in the retrospective thinking aloud, when he observed that he had not thought about his translation process before.

Another distinctive behaviour seen in Marcus was the frustration mentioned and the observed restlessness during and after the setback task. This behaviour seemed fitting for a non-buoyant participant. As mentioned in the retrospective interview, he wanted to perform well. Interestingly, his focus in deciding if it was going well seemed to lay on the task outcome and how quickly he was able to work.

### 4.1.2 Horatia

Horatia's high proficiency made it possible for her to rely heavily upon her Latin knowledge, and generally, few other strategies were observed. Horatia showed little metacognitive activity throughout the Latin Buoyancy Task, but when she did, it always led to meaningful subsequent action. Particularly interesting is that there is one sentence in which she let go of this very morpho-syntactical centered approach, only to return to it in the next sentence. It seemed she shortly was affected by something, but quickly picked her successful method up again, when starting afresh.

When Horatia was unsure about something, such as what type of word *traiectus* was, she tried different approaches to solve it, but then without losing her composure decided upon an answer, even though she was not sure about it and moved on without looking back. This unaffected manner and deciding to move on seemed in line with her high academic buoyancy.

Interestingly, in the case of the setback task, Horatia felt she still scored well. This might imply that she did not feel excessively challenged by the setback task, leading her to not have to display her ability to bounce back as much. However, she did say that she felt confused after the setback task. This *did* seem to imply she was aware of the difficulty, as feeling confused did not seem in line with saying she was expecting a good score. Possibly she was basing this score on what she was used to score and not on how she was feeling after the task. Particularly as she mentioned not understanding the meanings of the pre-setback task sentences, despite scoring well. It is possible that Horatia's analytical approach led her to score well generally on translation tasks, despite not understanding the content of the texts. Particularly in the lower years of Latin the content of the texts is often less important than the correct representation of the grammar, and this is even the case to some extent in the higher years (Kroon & Sluiter, 2010).

### 4.1.3 Marcus

Low proficient and high buoyant Marcus generally used good problem-solving strategies. He for example used knowledge of the world and tried to lean on text comprehension (Eikeboom, 1970; Florian, 2015; Karten, 2015). However, his lack of Latin morpho-syntactic knowledge meant he was only able to apply these strategies to a certain extent. Sometimes this even led to hypercorrection, thinking that his translation was too simple for a Latin translation task. He explained that he felt it was too much effort to learn the Latin grammar by heart for what it would yield. This unwillingness to learn things by heart if it lacks usefulness in the eyes of the student is a typical trait of gifted students (Van Gerven, 2002), moreover, generally when a person places much effort into something and this is not rewarded sufficiently they lose motivation to do so (e.g. Byrka, et al., 2021). Seeing his metacognitive abilities and the quality of the strategies he employed, it would be interesting to see how well he could translate if he could to do so using an overview of all the grammatical elements besides the supplied word list.

Marcus was generally displaying strategies indicative of academic buoyancy: he kept focusing on that it was not a problem if he did not know something and he might have learnt something in the meantime. His approach was generally fluid and he used different strategies in no particular order throughout the task. Trial and error seemed leading in Marcus' translation process.

Marcus commented on the effect of the setting on his approach to the task. He said he found it less stressful than he would have if it had been a proper exam setting. He was aware of his lack of knowledge and in this case felt more at liberty to try things. This lack of stress led him to remember Latin knowledge in particularly the setback and post-setback task that he had not used in the pre-setback task. He even reflected upon this during the task 'why did I not think about that earlier?'[1], indicating he was aware of what he was doing and had been doing throughout the task. This observation did not seem to hinder him or frustrate him, he just continued translating the current task. This strategy use seemed to fit with his high academic buoyancy as seen in both the Anagram Task as well as the Latin Buoyancy Task (Fried and Chapman, 2012; Meneghel, et al., 2019).

[1] Waarom dacht ik hier niet eerder aan?

#### 4.1.4 Quintina

Low proficient and low buoyant Quintina's metacognitive awareness was strong, but she often lacked the cognitive knowledge to act upon this. At first, she did not even try to act when she knew something was wrong. Her accuracy increased, due to her spending more time on problems and not automatically giving up by just going for something. Quintina was aware of her low proficiency and commented on it being a shame she did not know certain things, as she was aware of how that knowledge might help her in translating.

Quintina's translation process seemed to evolve and improve as she went. She tried new things throughout the task and if she felt it was helpful, such as writing the words above the text, she maintained doing that for the rest of the task. This adapting seemed to lead her to improving her score in the post-setback task. Particularly noteworthy is that Quintina was one of only two students who spent longer on the post-setback task than on the pre-setback task. This also seemed to coincide with her practicing more different strategies and having developed a method as she went along.

This adaptability and improvement shown by Quintina is unexpected as her academic buoyancy was low in the Anagram Task. However, this can be explained by her comment in the retrospective thinking aloud: she said the one-on-one setting with the researcher and the fact she had to perform the task thinking aloud led her to persist when she normally would not have persisted. This implies that academic buoyancy might not be domain specific, but can be affected by the setting of the task.

### *4.2 Comparing and contrasting: proficiency*

In the case of proficiency, we particularly expected differences in that [1] the high proficient participants would most notice the increase in difficulty during the setback task and [2] display more meaningful metacognitive activities, whereas the low proficient participants would be likely to use irrelevant information when monitoring or problem solving. Moreover, we expected [3] bottom-up strategies to be more prolific in the low proficient participants compared to the high proficient ones. When comparing the low and high proficient participants the following can be concluded:

1. All participants noticed the increase in difficulty, leading to changes in their translation processes. The low proficiency participants spent longer trying to solve the setback task, whereas the high proficient participants decreased their monitoring behaviour during and after.

2. The low proficient participants indeed showed incongruity between what they metacognitively observed and their subsequent actions.
3. High proficient participants leant mainly on an analytic method, unexpectedly showing mainly bottom-up strategies, whereas the low proficient leant on other strategies and tended towards trial and error strategies.

A wholly unexpected result was that the low proficiency participants both mentioned persisting longer and trying different (new) things due to the setting; they seemed to incur an effect of the study itself during the task. This was not the case for the high proficient participants. They said to follow their usual method. The one-on-one situation seemed to have led the low proficient participants to either not giving up (Quintina) or feeling less stress (Marcus) during the task. Whatever the case, the methodology of the current study seemed to have had an unintentional learning-on-the-spot effect on the low proficient participants.

In the following paragraphs we first compare Anthony's and Horatia's translation process. Then we similarly discuss Marcus's and Quintina's translation process and contrast this with Anthony's and Horatia's.

#### 4.2.1 High proficiency strategies: Anthony and Horatia

Anthony and Horatia both seemed to particularly monitor their translation when they were feeling uncertain about something. When monitoring they would approach the problem by using relevant cognitive knowledge. This knowledge was usually based on grammatical Latin knowledge. The use of Latin knowledge seemed central to the translation process of the high proficiency students. Under normal circumstances both students were analytical and focused upon determining the grammatical elements of the sentence. They both relied on bottom-up strategies to come to correct translations.

It seems safe to say that Anthony noticed an increase in difficulty in the setback task and then a decrease into the post-setback task: his perceived scores fluctuated from 50% to 45% and then up to 65%. Anthony also specifically mentioned in the post-interview that he noticed it was not going well in the second task. Horatia was more notable in this regard, as her perceived score for the setback task was higher than before the pre-setback task. A possible explanation for this is that she related her score to the effort she put into translating these sentences or that she relied upon her knowing how well she usually did in translation tasks and thus expected a similar score. Another possibility is that she did find the setback task difficult but did not want to admit this explicitly. This might indicate a socially desirable answer, which is not uncommon in adolescents (e.g. Malmberg, Hall & Martin, 2013). Support for her actually finding the setback task difficult might be found in her indicating feeling confused on the Emoji Chart directly after translating the setback task. Whatever her actual feelings on the difficulty of the setback task, one thing can be said: her task process altered after the setback task, indicating an effect.

This was similarly the case for Anthony: both our high proficient participants spent about the same time on the pre-setback task and setback task. After the setback task, they each spent a lot less time on the task. They were, thus, both affected by the setback task regarding their time on task, despite a difference in academic buoyancy. Both participants spent less time on monitoring and were content with their first attempt of a translation, whereas prior to the setback task both participants would continuously monitor their outcome. It thus seems that high proficient participants, despite their different academic buoyancy levels, monitored less well when being offered a simpler task after a challenge.

The decrease in difficulty seemed to have led to them becoming overconfident and, thus, they felt less need to monitor and evaluate their translation process. This idea of overconfidence seems to be supported by the perceived scores the students gave their translation of the third set: Anthony's perceived score was 65% compared to the 16,67% in reality, Horatia's score was only 50%, but she thought she had scored 85%. Both students grossly overestimated the correctness of their translation. There are studies that have found correlations between high cognitive skills and confidence levels (e.g., Białek & Domurat, 2017). Further research specifically looking at the effect of a (gifted) students' confidence in their translation on their subsequent metacognitive activity might be provide fruitful in further explaining the distinct decrease in use of metacognition after the undoable task. High ability students might benefit from actively being taught how to monitor and evaluate their work, particularly when they felt they performed well and thus monitoring and evaluation might be unnecessary.

#### 4.2.2 Low Proficiency strategies: Marcus and Quintina

Both Marcus's and Quintina's perceived scores on the setback task decreased compared to the pre-setback task, indicating they were aware of a change in difficulty. However, they also both mentioned finding the pre-setback task difficult. There is a change in their translation process, with both Marcus and Quintina having tried different things. This was also reflected in the fact that both students spent significantly longer on the setback task than on the pre-setback task. However, both participants admitted that they would not have tried for so long if they had been translating in a normal setting, so there seems to be an effect of the methodology in the case of the low proficient participants.

Marcus and Quintina often lacked the precise Latin knowledge they needed to solve problems, but would say what they did know, despite its often irrelevance to the problem at hand. Both participants explicitly mentioned not possessing of the needed Latin knowledge to complete such a task well. Interestingly, they both discovered they knew more than they thought as the task progressed. The incongruity between their monitoring observations and following actions often arose from them not having more relevant Latin knowledge at their disposal.

Both participants did not show a predetermined approach but adapted their method to what knowledge they had available to them. Usually this was knowledge of the Roman or general world, but towards the end of the task they made increased use of Latin knowledge.

#### 4.2.3 Comparing and contrasting high and low proficiency strategies

All four participants, with the possible exception of Horatia, noticed the increased difficulty in the setback task. All participants adapted their method: the low proficient participants spent longer trying to come to an acceptable translation and sometimes stumbled upon useful strategies that they used in the post-setback task. There seemed to have been a learning-on-the-spot effect of method in the case of the low proficient participants.

Studies aimed at determining whether thinking aloud affects the participants' process have generally found this not to be the case when performing complex tasks. Thinking-aloud is, however, to be a method that demands more time (e.g. Ericsson & Simon, 1993; Leow & Morgan-Short, 2004; Park, Korbach & Brünken, 2020). Furthermore, students must put in more mental effort when thinking aloud (Park, Korbach & Brünken, 2020). Possibly the thinking aloud invited the low proficiency participants to put more effort into the task, than they would have otherwise. Based on the insights Marcus and Quintina came to and then continued to apply, it might be advisable for teachers to spend some one-on-one time with their low proficient students and whilst doing so ask their students to translate thinking aloud.

The high proficient participants were indeed effective in their use of metacognition, but this was partly due to their ability to produce a larger range of Latin knowledge: the high proficient participants had more knowledge to their disposal, so the chance of them knowing the relevant information was larger. There were multiple examples of the low proficient participants knowing what would be a useful step, but that they were unable to follow through on that procedural knowledge due to a lack of declarative knowledge. To better determine whether a student is low proficient at Latin due to a lack of specific Latin knowledge or due to a lack of metacognitive skills it might be interesting to let low proficient students perform short translation tasks with an overview of the morphological paradigms.

Unexpectedly, the high proficient participants displayed mainly bottom-up strategies, whereas the low proficient leant on other (more creative) strategies and developed a method as they went. This is opposite to what Boyd (2018) and Luger (2018) found in their studies. Two explanations might be given for this contradiction. First, previous translation studies focused on participants who were translating continuous texts not separate sentences. Continuous texts might provide students with more context for text comprehension and therefore proficient students might be more inclined to use top-down strategies than when translating separate sentences, such as in our study.

Another explanation might be to do with the age of our target audience. Our participants are younger than those from the studies of Boyd (2018) and Luger (2018). According to Veenman & Spaans (2005), students are generally not able to transfer metacognitive skills from one domain to another until they are 15 years old. Our participants were generally younger. Possibly they have yet to develop the necessary skills to coordinate top-down strategies.

### *4.3 Comparing and contrasting: academic buoyancy*

In the case of academic buoyancy, we particularly expected differences in that [1] high buoyancy participants would be more flexible in their process particularly when experiencing difficulties and that [2] they would experience fewer negative emotions such as frustration compared to the low buoyancy participants. Moreover, we expected that [3] avoidance strategies would be observed in the low buoyant participants, but not in the high buoyant participants.

When comparing the low and high buoyant participants the following can be concluded:

1. The high buoyant participants were indeed more flexible in using different strategies to approach the same problem.
2. Negative emotions were particularly seen in the low buoyant participants.
3. Avoidance was particularly seen in the low buoyant participants.

The low proficient participants seemed influenced by the setting of the task. This difference in their behaviour must be taken into consideration when discussing the low academic buoyancy results. In section 2.1 we presented that finding a participant to represent the low buoyant and low proficient student was less straightforward than for the other profiles. Despite the large decrease in the Anagram Task, Quintina's score improved on the Latin Buoyancy Task after the setback. Quintina said in the retrospective thinking aloud that she persisted more than she normally would. Persistence on its own would seem more fitting in high buoyant participants (Martin & Marsh, 2008). However, the addition that it was 'more than normal' is of particular note.

Quintina's reflection on her persisting was confirmed by our observations. At the beginning of the task she did not act when coming across problems, she only observed that there was a problem, which seems in line with being not so buoyant and thus with the results from the Anagram task. Later, she indeed persisted more. Thus, the process shown during the pre-setback task seems to be more representative of Quintina's normal task process. Therefore, these sentences were taken as the base for conclusions on low buoyancy processes.

In the following paragraphs we first discuss the high academic buoyant participants and their task process and displayed strategies, followed by that of the low academic buoyant participants before cross-referencing them.

### 4.3.1 High academic buoyancy strategies: Horatia and Marcus

Both Horatia and Marcus were flexible in trying to find different solutions when they were stuck. They each showed proof of believing they were able to learn from difficulties. According to Aslam and Ali (2017), students who are motivated by failure to try another strategy and put in more effort are more likely to be academically buoyant. This seems the case for Horatia and Marcus: Horatia said she believed she would be better at the second task as she had already performed a similar task before, whereas Marcus kept focusing on that it did not matter if he did not know things and was aware of developing skills as he went. This approach to difficulties is reminiscent of how students with a growth mindset are thought to view the relationship between learning and being challenged (e.g. Yeager & Dweck, 2012). In Chapters 4 and 5 we also found a relationship between students having an growth mindset and being buoyant.

Regarding negative emotions and avoidance strategies, few were observed throughout the Latin Buoyancy Task in either of the academic buoyant student. When negativity was observed, this was relatively late in the tasks compared to the low buoyant participants. Moreover, Marcus also displayed positive emotions in the setback task. Thus, affectively it seems that Horatia and Marcus demonstrated strategies as expected based on the descriptions of academic buoyancy by Martin and Marsh (2008).

### 4.3.2 Low academic buoyancy strategies: Anthony and Quintina

To minimize interference from the effect that the task setting had on Quintina, we only include the pre-setback task, i.e. the task of which the aim was to gage the participants' translation process under normal circumstances. Due to her low translation proficiency, she encountered problems within that task, and investigating how she reacts to these setbacks should still provide information on how her low buoyancy affected her strategies.

Anthony and Quintina were not flexible in their approach to problem-solving and on occasion demonstrated avoidance strategies, similar to what Newland (2016) saw in his translation study. After running into problems, both students displayed negative emotions. Anthony, in particular became frustrated as attested by for example his frequent ticking on the desk with his pen (Farley, Risko & Kingstone, 2013). Combined with his high proficiency this can be explained by him noticing the increase in difficulty (Martin and Marsh, 2008), and him wanting to do well causing a fear of failure (Goossens, 2004; Kieboom, 2015; Newland, 2016). In the case of Quintina, she was critical of her not knowing things and seemed to put herself down. She seemed not to believe in her being able to successfully complete the task and therefore did not fully commit to the task. The belief a student holds regarding their capability to complete a task has been given as the definition of self-efficacy and can influence decision making when challenged (Bala, Kaur & Singh, 2017).

### 4.3.3 Cross-referencing Anthony and Horatia

In both high proficient participants, we saw that the metacognitive behaviour nearly totally disappeared after the setback task, which seemed to have affected their scores to a certain degree. Both participants' scores decreased, Horatia's a little and Anthony's significantly. Horatia still relied on grammar during the translation of the third set, whereas Anthony discontinued to refer to any grammatical knowledge; he focussed on word meanings. It seems that the low buoyant participant's process was indeed affected the most. As expected, the participant with low buoyancy and a high proficiency was most affected during the setback task; Anthony was aware of the difficulty, but affectively unable to deal with it, which led to him being less flexible when trying to solve problems and more negative.

According to Martin (2013) low academic buoyancy is related to achievement anxiety, poor isolated grades and temporary lapses in motivation. This was indeed seen in the behaviour of Anthony and not in Horatia: in the interview he mentioned wanting to do well and becoming frustrated when he did not (achievement anxiety), his accuracy decreased after the setback task (i.e. poor isolated grades) and he gave up toward the end (i.e. temporary lapse in motivation). Meanwhile, among other things persistence and confidence are correlated with high academic buoyancy (Martin & Marsh, 2006; Martin, et al., 2010). The task thus seems sufficient in discerning differences in the translation processes of high proficient students with different buoyancy levels.

## 5 CONCLUSION

The aim of this Chapter was to create four strategy profiles for students by determining behaviour specific to students depending on their academic buoyancy and Latin translation proficiency. When participants translated the Latin sentences and ran into a larger problem, there was a prolonged effect of the difficulty on their translation process. The exact effect of the difficulty increase was dependent on the participants' level of Latin translation proficiency and on their level of academic buoyancy.

To create the profiles, we created a quadrant containing the participants' proficiency on one axis and buoyancy on the other. Generally, by making use of a detailed qualitative analysis, we found that the participants either included new cognitive strategies or emotions not seen in the pre-setback task, or discarded strategies when confronted with the setback task, and were thus affected by the increase in difficulty.

The increase in difficulty led to the following effects cognitively, metacognitively and affectively:

1. *Cognitively* the high proficient participants surprisingly made use of mainly bottom-up strategies. Moreover, the low proficient participants tried to apply more cognitive knowledge when the difficulty increased, with the research setting seeming to influence the students' translation process. This led us to conclude that [1] top-down strategies which include text comprension might already be taught to lower year students, particularly those who have mastered the grammar in theory to improve translation proficiency and that [2] low proficient students might improve their skills by translating whilst thinking aloud.

2. *Metacognitively* the high proficient participants were more aware of difficulty changes and more able to use effective strategies as reaction to their monitoring. The metacognitive awareness of the high proficient participants in particular lessened after the setback task. These students might thus profit from being actively taught how to check if their perception of the task being easy is indeed the case or how to evaluate a translation when they are not expecting to find any problems.
3. *Affectively* the low academic buoyant participants were indeed more prone to avoidance strategies, and they were also less able to not become negative about their abilities or frustrated when challenged. This was particularly the case when the participant was high proficient with a fear of failure (Anthony). Seeing how others deal with challenges and regulate their emotions in such cases might help them develop more coping skills and subsequently become more academically buoyant.

Overall, the differences between participants did not lay as much in the quality or quantity of cognitive, metacognitive and affective strategies as independent strategies rather than in the quality of the interaction between these strategies. An example of this was that the low proficient participants monitored problems correctly, but often did not follow up on these observations with cognitive strategies related to the problem at hand. The high proficient participants did show more coherence between their displayed subsequent metacognitive and cognitive strategies.

# CHAPTER 7

# GENERAL DISCUSSION

The purpose of this dissertation was to explore how a lack of academic buoyancy in gifted students is reflected in their task process after being confronted with a setback. We also investigated to what extent the effect of the setback task on the participants' translation accuracy and process was moderated by learner variables. We did so in a Dutch context using a task-based measure. The current, final, chapter consists of six sections. In the first section, we summarize our main findings regarding academic buoyancy and its moderators. In the second section, we discuss the validity of the studies presented in the previous chapters. This is followed by a third section dedicated to practical implications for Latin teachers and students, that arise from our findings. In the fourth section, we propose directions for further research. In the fifth section, we follow up on the participants. Finally, we end this chapter and dissertation with our concluding reflections.

# 1 SUMMARY & MAIN FINDINGS

In this dissertation, we explored academic buoyancy in a Dutch context through a task-based measure, the Latin Buoyancy Task. This task was designed to include a setback. It was presented in Chapter 1 alongside the design of the other instruments used in this study. The first testing of the Latin Buoyancy Task indicated that it was indeed possible to study academic buoyancy through a task-based measure. As explained in Chapter 1, we preferred this type of measurement over a self-report questionnaire.

In Chapter 2 we described how the Latin buoyancy task would be used to gather data by asking the participants to perform the task whilst thinking aloud. Besides concurrent thinking aloud, we also included retrospective thinking aloud as a method of data-collection. In the second chapter, we also presented how we made use of extreme sampling to select participants that were either expected to be particularly academically buoyant or non-buoyant. This led to the inclusion of 16 participants to perform the Latin Buoyancy Task whilst thinking-aloud.

In Chapter 3 we first presented which decisions led to the creation of the verbal transcripts. We also demonstrated why and how we created two coding schemes. One coding scheme focused on the content of the participants' utterances, whilst the second focused on context bound interpretations of the utterances, providing deeper insight in the affective component of certain reactions. Finally, the coding procedures were also included in Chapter 3.

In the remaining chapters we presented the results from the data analyses. The analyses included three learner variables (translation proficiency, mindset preference and frustration tolerance). In the course of chapters 4-6, we asked to what extent and in which manner these learner variables moderated the effect of the setback on the participants' translation process, regarding their cognitive, metacognitive and affective strategies. As the chapters progressed, we further zoomed in on the verbal transcripts in our analyses.

The first of our analyses was quantitative, presented in Chapter 4. We quantitatively compared and contrasted the 16 participants' translation accuracy of the pre-setback task to that of the post-setback task to determine whether there was an effect of the setback task on the quality of their translations. We found that there generally was a significant effect of the setback on the quality: it was lower for the post-setback task than for the pre-setback task. We also asked whether there were indications of the participants' having altered their translation process. We found that participants generally displayed less implicit metacognition (i.e. switched less between thinking activities), persevered less (i.e. spent less time on task) and were less gritty (i.e. switched quicker between thinking activities) in the post-setback task compared to before. Finally, we investigated whether the learner variables moderated the effects of the setback task. We found that the effects of the setback were larger when the participant was more proficient or was more inclined towards a fixed mindset. The results regarding frustration tolerance were less clear than translation proficiency and mindset preference. The results led us to exclude frustration tolerance in further analyses.

In Chapter 5 we qualitatively analysed the verbal transcripts of the part in the task in which the participants had to deal with the setback, translating three impossible Latin sentences. For this study, we included only the 16 participants that were among the highest and lowest proficient participants or had the most outspoken mindset preferences. We aimed to answer whether the effects as seen in Chapter 4 were already visible during the setback task itself, and they were. Moreover, we aimed at answering how translation proficiency and mindset preference accounted for differences between the participants' display of metacognitive and affective strategies throughout the setback task. We found that proficiency and mindset independently of each other had effects on the participants during the setback task. The analysis led to further questions particularly the case of the proficiency participants. These questions might be answered by qualitatively analysing the participants translation process throughout the whole Latin Buoyancy Task.

In Chapter 6 we, therefore, qualitatively delved further into the verbal transcripts. To strengthen the analysis, we also included the retrospective interviews, and all the other data available to us regarding these participants. In Chapter 6 we qualitatively analysed all data of four participants. Each participant represented one part of the quadrant that included the 'extent of academic buoyancy' on one axis and the 'extent of proficiency' on the other. We concluded that low and high proficient participants demonstrated differences regarding their cognitive, metacognitive and affective strategies, depending on whether they were high or low academically buoyant.

Looking at the dissertation as a whole, our main findings can be summarized as follows:

1. Academic buoyancy can be studied via a task-based measure.
   Previous research on academic buoyancy has been based on questionnaires (e.g. Malmberg, Hall & Marin, 2013; Kim & Han, 2014; Comerford, Batteson & Tormey, 2015; Smith 2015; Jahedizadeh, Ghonsooly & Ghanizadeh, 2019). Perseverance and grit have previously been operationalised in such a

manner that they could be studied under 'lab conditions' (Lucas, et al., 2015; Gerhards & Gravert, 2020). Our results indicate that a temporary setback can also be isolated in a task and, therefore, our knowledge of academic buoyancy can be expanded by data collected in other ways than questionnaires.

2. The heterogeneity of Latin texts necessitates academic buoyancy.
   The results also indicated that the heterogeneity of Latin texts necessitates the translator to be somewhat academically buoyant. We found that some of our participants did bounce back after the setback, however others did not. For some of the low buoyant participants, the decrease in their accuracy was particularly significant.
3. The effect of the setback on the participants' translation accuracy is related to an effect of the setback on the participants' translation process and can be a prolonged effect.
   Our results indicate that the translation accuracy of the participants significantly decreased after the setback, despite the complexity of the pre- and post-setback task being comparable. However, not only their accuracy was altered by the setback. For example, by the end of the setback task, metacognitive strategies were barely demonstrated, even when they were present in the pre-setback task. Furthermore, we observerd changes to their translation processes. Some participants maintained their altered translation process as observed in the setback in the post-setback task and thus did not bounce back to their pre-setback translation process. In most cases, except in some low proficient participants, the effect of the altered process was negative regarding their translation accuracy.
   In the case of the high proficient participants, we found that in the post-setback task, they did not demonstrate the same negative emotions as they did in the setback task. They seemed to somewhat overcome these emotions. However, they did become overconfident and worked more sloppily compared to the pre-setback task. It is this change that seems to explain why their accuracy declined after the setback and what hindered them from bouncing back.
4. The participants' level of academic buoyancy in the Latin Buoyancy Task was influenced by their translation proficiency and/or mindset preference.
   Students who are excellent in Latin, might suffer the most from a difficult subtask. The most proficient translators (i.e. scoring a perfect score on the pre-setback task) adapted their translation process the most during the setback task, particularly indicated by their diminished use of metacognition

and them relatively quickly demonstrating negative affection. We suspect this happened because due to their high proficiency, they were more aware of the increased difficulty during the setback task. They continued using this 'new' translation process in the post-setback task, resulting in a significant decline of accuracy. Students who are good but not excellent might also suffer, but to a lesser extent. Some of the other participants that scored near perfect scores on the pre-setback did bounce back to their pre-setback translation process after the setback task and still their accuracy also declined. However, their decline in accuracy was not significant. Thus, when the participant was a high proficient translator, they bounced back less.

For mindset we similarly saw that when a participant was more inclined towards a fixed mindset, they adapted their translation process more than those inclined towards a fixed mindset during the setback task. The extensive qualitative analysis suggests that they continued using their 'new' translation process in the post-setback task.

Thus, both a high translation proficiency and/or a fixed mindset increases the possibility of the participant being less academically buoyant when translating a heterogenous Latin text.

5. The interaction between on the one hand affective strategies and on the other cognitive and/or metacognitive strategies played a major role in the extent that the participants bounced back from the setback task.

   Participants who had reacted more emotionally during the setback task, scored worse on the post-setback task compared to those who had not. We found that when a participant's academic buoyancy was low, the setback task had triggered the display of affective strategies more compared to the academically buoyant participants. Not only was this the case during the setback, but we saw a prolonged effect of this in the post-setback task. Moreover, their affective strategies were less helpful than those of more buoyant participants. The lack of helpful affective strategies led to a decrease in the quality of the cognitive and metacognitive strategies during and after the setback. We concluded that the low academically buoyant participants were unable to regulate their affective strategies helpfully during the setback and this affected their post-task translation process as well.

   This effect was strengthened when the low buoyant participant was also high proficient and, thus, more aware of the setback. These participants seemed to lose their confidence and motivation during the setback task, after first displaying an increase in effort. We concluded that the relief (i.e. an emotion) of the easier ensuing post-setback task continued to cloud their cognitive and metacognitive strategies: they became overconfident. Therefore, they felt they did not need to monitor and evaluate their cognitive activity as much, which in turn led to a much less accurate translation than pre-setback (see Chapter 6).

In Chapter 5 we also saw indications that affective strategies particularly influenced the translation process of the participants with a fixed mindset preference. The lack of helpful coping mechanisms (c.f. Hong, et al., 1999; Cook & Artino, 2016) led participants with a fixed mindset preference to avoid problems during the setback task.
In sum, affective strategies and their effect on cognitive and metacognitive activities played a major role in the extent that the participants bounced back from the setback task. This effect was in turn moderated by the participants' translation proficiency and mindset preference.

## 2 VALIDITY

In the current section, we put potential validity issues with the study forward. We do so in three parts. We start with discussing the validity of the main measurement of this study: the Latin Buoyancy Task. We then assess the methodology including participant selection, the think-aloud methodology and our analysis of the verbal transcripts, including the translation process variables. Finally, we address possible validity issues with how we determined the learner variables translation proficiency, mindset and frustration tolerance.

### *2.1 The Latin Buoyancy Task*

The aim of the Latin Buoyancy Task was to isolate an academic setback and enable an investigation of academic buoyancy. In validating the instrument, we must first ask whether we succeeded in creating and isolating a setback. To validate the instrument, we asked the participants what they felt their translations of the individual tasks were worth. In the retrospective thinking aloud we invited the participants to elaborate on their experiences via stimulated recall. First we address the length of the Latin Buoyancy Task and the inclusion of unrelated sentences instead of an ongoing text. Next, we discuss how the participants' subjective scoring helps validate the instrument. We then move on to the verbal transcripts, before comparing our instrument to task-based measures of psychological phenomena used in other studies. Finally, we discuss to what extent the Latin Buoyancy Task indeed measured academic buoyancy.

A possible critique of the Latin Buoyancy Task was that each of the task only consisted of three sentences. The small number of items in the pre- and post-setback tasks made it difficult to guarantee their internal consistency. However, translation is a demanding task, which was made more cognitively demanding by letting the participants perform it whilst thinking aloud. Intensifying the think aloud sessions by lengthening the tasks would likely have a detrimental effect on the quality of the translations, particularly on the post-setback. This in turn would cloud the effects of the setback task. As it was, the think aloud sessions lasted on average around an hour. Lengthening the task, albeit likely improving the internal consistency of the pre- and post-setback tasks, would overall not improve the validity of the instrument.

A second possible question that could be asked regarding the Latin Buoyancy Task regards the somewhat artificial nature of the task. We did not make use of sentences that formed a continuous story. Therefore, translation skills related to text comprehension were only minimally included in the task on a sentence-based level. If we had conducted the study amongst older students, this might have been a relevant issue with the task. In the upper years, the texts students translate from Latin are continuous texts, usually in their authentic Latin form. In the lower years, the texts the students translate are so-called schoolbook Latin: texts written by schoolbook publishers. These texts tend to be unnatural in that they focus on the newly learnt grammatical features. Students would thus not be used to texts that are not specifically built up around one or two grammatical features. Furthermore, in the first three years, students' regularly practice and consolidate new grammatical features via exercises that involve translating unconnected sentences. Hence, the Latin Buoyancy Task was similar to tasks they regularly perform during Latin lessons and therefore, the artificial nature of the Latin Buoyancy task was not problematic.

Table 7.1 presents the scores the participants gave their translations. The participants were asked to score their translation out of ten points. This is the common grading scale in the Netherlands, with 10 standing for a perfect translation and anything higher than a 5.4 qualifying as a pass. From Table 7.1 it becomes clear that all the participants, bar Livia, gave their translations of the setback task the lowest score of the three tasks. In other words, they felt that their performance on this task was at its lowest. Besides Diana and Livia, all scores for the setback task were failing scores and Diana's mark was a barely passing mark. The lower scores indicate that the participants noticed the increase in difficulty. Nona even refused to give her translation for the setback task a mark as it went so badly. She explained why she gave herself a mark for the third task 'Because this one did go better than that previous one. I was like: luckily this one is not going even worse'[1]. The failing nature of their tasks can be interpreted as the participants being aware that they had not completed the task well. According to Martin & Marsh (e.g. 2008), difficult schoolwork and poor grades are both examples of academic setbacks that students have to deal with on a daily basis. Moreover, it is these daily setbacks that academic buoyancy refers to. These are indications that our task did indeed isolate a setback and therethrough measured academic buoyancy.

Further, the verbal transcripts provide other indications that the setback task was indeed experienced as a setback. During the setback task there were instances of participants specifically referring to finding the task difficult (e.g. Ennius: 'I find this second set difficult'[2]). Also, during the retrospective thinking aloud participants made comments that identify the particular difficulty of the setback task. Diana said she did not often expects to receive a failing grade, 'but this was a bit more difficult'[3]. The decrease in difficulty for the post-setback task was also noted. For example, looking back on the third task, Marcus said when asked why he gave himself a higher score for the post-setback task than the setback task 'It is more that with this I thought "okay this is correct", but with the other I had absolutely no idea what it was supposed to say'[4]. Octavia in turn said regarding the post-setback task: 'these sentences were shorter and that is easier'[5]. The participants' concurrent and retrospective thinking both indicated their experience of an increase and decrease in difficulty in the setback task and the post-setback task respectively. Thus, the instrument is a measurement of academic buoyancy.

This is further strengthened when we compare our measure to an instrument created as a task-based measure of grit (Lucas, et al. (2015). These researchers included different impossible tasks to measure grit in their participants and found positive results in identifying gritty behaviour via a task-based measure. They only focused on specifically identifying perseverance and grit during the difficult tasks, not on general task process, or whether there were prolonged effects of the challenge. However, that they were able to isolate grit within a task-based measure is indicative that it could be extended to academic buoyancy, a larger phenomenon which includes grit.

---

[1] *Omdat deze wel beter ging dan die vorige had ik wel iet s van gelukkig gaat deze niet nog slechter.*

[2] *Deze set 2 vind ik wel moeilijk.*

[3] *Maar dit was wat moeilijker.*

[4] *Het is meer dat ik bij deze had oké, de dit klopt wel, terwijl ik bij die andere echt geen idee had wat er moest staan.*

[5] *De zinnen waren korter en dat is makkelijker.*

*Table 7.1 The participants' subjective scores per task*

| | Pre-setback | Setback | Post-setback |
|---|---|---|---|
| Anthony | 6.5 | 4.5 | 5 |
| Bella | 5 | 4 | 5 |
| Claudia | 5 | 4.5 | 6.5 |
| Diana | 7.5 | 5.5 | 6 |
| Ennius | 5.5 | 4 | 8 |
| Flavia | 6.5 | 4 | 4 |
| Gaius | | 1 | 8 |
| Horatia | 6 | 5 | 8 |
| Julia | 7 | 5 | 8 |
| Livia | 7.8 | 8 | 8.5 |
| Marcus | 4 | 1 | 4 |
| Nona | 6.5 | | 7.5 |
| Octavia | 4 | 3 | 5 |
| Phaedra | 6.3 | 4 | 7.5 |
| Quintina | 4 | 2 | 3 |
| Rufus | 6.5 | 5.9 | 6.5 |
| *M* | 5.87 | 4.09 | 6.28 |
| *SD* | 1.24 | 1.84 | 1.75 |

*Note. The scores were based on a 1-10 scale, with a 10 standing for a perfect translation.*

Some might argue that we could have determined the extent academic buoyancy on the basis of the Academic Buoyancy Scale (Martin & Marsh, 2008) instead of via a task-based measure. However, academic buoyancy was the central topic of our study and a questionnaire might still not have provided us the participants' extent of academic buoyancy. Participants answering the questionnaire with socially desirability would have significantly clouded our results. Adolescents are known to engage in social desirability in self-report (Fan et al., 2006). For example, in the case of perfectionism it has been demonstrated that adolescent participants structurally report to more perfectionistic behaviour than they exhibit in reality (Stoeber & Hotham, 2013). Moreover, there are concerns that due to distorted self-perceptions, self-reports are often inflated (c.f. McDonald, 2008). Thus, validity issues might have arisen by using a self-report questionnaire as the principal measurement of academic buoyancy. For us, this was of particular importance as academic buoyancy had not previously been investigated in a Dutch context. Thus, there were also possible issues with cultural bias (c.f. Taber, 2014).

In addition, a questionnaire filled in by the teacher regarding the participants' academic buoyancy was also no possibility. Verrier, Johnson and Reidy (2018) disproved the validity of such an instrument for academic buoyancy. They concluded that teachers are likely to be influenced in their assessment of academic buoyancy by other learner variables and do not actually report on students' academic buoyancy. Thus, teacher report was also not an option.

Finally, this dissertation was predominantly centred in the educational domain. By creating a task-based measure, our observations would be closely related to what teachers observe during their lessons, making the descriptions of task behaviour more recognisable to teachers. Therefore, the impact of the findings might be easier to transfer to the classroom.

### *2.2 Methodology*

Our study had a 'within subjects' design, meaning that all participants partook in the pre-setback, setback and post-setback tasks. The pre- and post-setback tasks were the same for all participants, for reasons explained in Chapter 1. The main advantages of a 'within subjects' design is that the variation between participants is minimized and research observations are more accurate (e.g. Charness, Gneezy & Kuhn, 2012). The disadvantage of the method is that there is a risk of 'carry over' effects, or, in other words, because the participants repeat the measure over time, they can learn as the tasks progress (McBurney & White, 2004). In our design that risk is minimalised. First, translating Latin into Dutch is a task that the participants perform multiple times a week. It seems unlikely that they will learn significantly more during the task itself. Secondly, the tasks were administered directly after each other, meaning that there was no time in between in which participants might learn new skills.

One might say that our results did suffer from a 'carry over' effect. For, Diana, our low buoyant and low proficient participant from Chapter 6, did explain that her results were affected; she did not give up and therefore her low academic buoyancy did not manifest itself as expected based on her results from the Anagram Task: her accuracy actually improved in the post-setback task. However, we propose that this was not a 'carry-over' effect of the 'within subjects' design, but an effect of the thinking aloud. After all, Diana was already well acquainted with translating. We elaborate on this in 2.2.2. First, however, we discuss the validity of the participants' selection. We also address the analysis of the data, first the quantitative analysis and then the qualitative analysis.

### 2.2.1 Participant selection

The Academic Buoyancy Scale was not used in the screening procedures for similar reasons. By selecting participants using a task that was similar in structure to the Latin Buoyancy Task, we presumed that we selected participants who were most relevant to our research question (e.g. Patton, 2002; Creswell, 2009; Palinkas et al., 2015).

Due to the lack of a generally accepted definition of giftedness, we operationalised giftedness by inviting participants from a *Begaafdheidsprofiel-school* with Latin in their curriculum. However, can we in hindsight say that our participants were gifted? As it turned out, 11 of the 16 participants had partaken in an IQ-test and had all scored a TIQ of >129.While an IQ-result is merely a momentary snapshot, for many studies related to giftedness, an IQ-result on its own is a satisfactory condition for participation in the study (e.g. Bennett-Rappell & Northcote, 2016; Makel et al., 2016). Our five other participants were successfully following their curriculum at the highest attainable level in the Netherlands. Therefore, we are of the opinion that our participants can at least be recognised as cognitively gifted.

Still, our participants remained limited to gifted participants that performed well at school, or at least have performed well enough at school previously to be able to include Latin in their curriculum. The scope of 'giftedness' is particularly broad, and one of the reasons that there is still no generally recognised definition of giftedness (e.g. Renzulli, 1986; Betts & Neihart, 1988; Ziegler & Heller, 2000; Gagné, 2003; Cramond, 2004; Renzulli, 2016; Papadopoulos, 2020). Generalizing results from studies including giftedness is problematic due to the broadness and individual nature of giftedness (LIT). By only including Latin students as participants, we somewhat resolved the problem of broadness. Moreover, our participants can be seen as a representation of the students that Latin teachers find in their classrooms, making the study relevant to those teachers. Our results, therefore, can and must only be seen in the light of Dutch, cognitively gifted students who have included Latin in their curriculum.

This also tackles the possible issue of distortion that can be put forward as a risk of extreme sampling (Patton, 1990; 2002). We did not generalize our findings, therefore, there is no risk of distorted results as a result of extreme sampling. For more on this, see Chapter 2.

### 2.2.2 Think aloud method

For the thinking aloud sessions, we aimed at gathering valid think aloud data, following procedures described by Ericsson & Simon (1993), Van Someren, Barnard & Sandberg (1994) and Young (2005). We included both concurrent and retrospective thinking aloud. We conjectured that this combination strengthened the validity and reliability of the analysis of the data. For, the interpretations of the data from the concurrent thinking aloud could be checked and nuanced according to what the participants said during the retrospective thinking aloud. This was particularly helpful in the case of the affective codes. Anthony, for example, mentioned feeling frustrated in the retrospective thinking aloud, something we would only have otherwise been able to interpret from his physical behaviour during the concurrent thinking aloud. The retrospective thinking aloud provided us with a confirmation of Anthony feeling frustrated. The inclusion of both concurrent and retrospective thinking aloud thus added to our understanding and helped validate our interpretation of the verbal transcripts.

Reactivity is a concern to validity when using a think aloud research design (Russo, Johnson & Stephens, 1989; Zhang & Zhang, 2019). Reactivity refers to the possibility that the participants' processes are changed by the thinking aloud itself. Many studies have shown that regarding internal reactivity, participants do not substantively alter their task process when thinking aloud compared to when they perform the task silently (e.g. Fox, Ericsson & Best, 2011). However, we did find examples of participants who acknowledged in the retrospective thinking aloud that their translation processes were altered by the setting. Diana said: 'Yeah, it is quite tricky. Yeah, I did think, that if had done this by myself, that I would have tried and had looked up the answers and that I would have actually already disengaged'[1]. It thus seems that the thinking aloud and the research setting in general led some participants to try for longer than they would have and this led to them stumbling on new strategies to employ which they otherwise would not have.

This would have been a breach in validity, had we not had the retrospective elaborations of the participants. In the current design, however, the participants were able to explain why they altered their process. We can include this knowledge in our analyses and therefore, we assume that the thinking aloud was

---

[1] *Ja het is best lastig. Ja ik dacht ook wel, dat als ik dit zelf had gedaan, dan had ik het wel geprobeerd en was ik gaan kijken naar de antwoorden en was ik eigenlijk allang afgehaakt.*

still a reliable method for the study. In this case, the occurrence of reactivity can be interpreted as an indication for teachers of what might help low proficient translators improve: thinking aloud in the company of another person might help them persevere. This is reminiscent of findings by Gerhards & Gravert (2020), who found that working in peer groups can motivate students to persevere longer. Thus, we do not think that the reactivity of the thinking aloud with a researcher present did not affect the data in such a manner as to make it invalid.

One of the measures we took to minimize reactivity caused by the researcher was related to how the researcher reminded the participants to think aloud when they fell silent. When a participant paused thinking aloud, we made use of short prompts to avoid prompting effects (e.g. Ericsson & Simon, 1993; Veenman, 2011). A pause during translating whilst thinking aloud is often interpreted as an indication of the participant performing a complex cognitive activity (Kujamäki, 2015). However, it can also indicate other things, such as a distraction (O'Brien, 2006; Kujamäki, 2015). Interpreting the cause of the pause is near impossible for researchers (Schilperoord, 2001).

We included the word 'think' in our prompts (*"think* aloud"), and in hindsight we may have been inadvertently interpreting the cause of the pause as related to cognitive processes. Our results indicate that when faced with a setback, affection seems leading in the decision making, despite there being relatively few direct references to affective strategies. Asking what the participants were *feeling* during a pause might have produced more explicate references to affection and demonstrated more of what was causing the participant to display certain cognitive and metacognitive behaviour. We thus might have inadvertently caused reactivity by the wording of our prompts. Future thinking aloud studies might do well to not only ask participants what they are thinking but also what they are feeling or include neutral prompt such as 'yes' with a questioning intonation, when the research question includes affective strategies.

### *2.3 Analyses*

In Chapter 4 we quantitatively compared the translation accuracy and processes of all 16 participants. We found that, after the setback, the accuracy of the most proficient participants decreased significantly and that the accuracy of some of the least proficient participants increased. One might propose that this was not an effect of the setback, but related to the statistical phenomenon of regression towards the mean. This phenomenon entails that when a participant has produced an extreme score, such as found in our highest and lowest accurate translation scores, in a next measurement, the participant's score is likely to be less extreme and closer to the mean. However, particularly in the case of our high proficient participants, the accuracy scores went from extremely high before the setback to extremely low after the setback. Therefore, their scores were no closer to the mean. In the case of the low proficient participants, regression to the mean might be somewhat at play, but for the high proficient participants the scores were no less extreme. Therefore, we propose that the change in accuracy was not just an effect of regression towards the mean, but also of the setback task.

In Chapter 5, despite creating process bars for all 16 participants, we only qualitatively analysed the bars of the participants who were either the most or least proficient or were among the participants with the most outspoken mindset preference. In Chapter 6 we further delved into the verbal protocols of four participants and analysed them qualitatively. The quantitative results provided us with the proof of our hypothesis that some students do not quickly recover their translation accuracy after a difficult passage. One of the reasons why we expected Latin translation to be an apt domain to study academic buoyancy, was related to the heterogeneous nature of the translation tasks: syntactically simple and extremely complex sentences may alternate each other somewhat randomly. Thus far in the study, the effect of the heterogeneous complexity of Latin texts had not yet been demonstrated. Via quantitatively comparing and contrasting the translation accuracy of the pre- and post-setback tasks, we demonstrated that the heterogeneous nature of the texts indeed may pose a problem for some students: not all our participants bounced back to their prior level after the setback, despite the complexity decreasing. This affirmed the necessity of further examining the effect that a difficult passage has on students.

The qualitative studies not only gave us more understanding of how and why the participants adapted their translation process, but also confirmed the findings and used measurements from the quantitative analysis. In the quantitative analysis we included three variables to operationalise task process: implicit metacognition, perseverance and grit. We measured these variables through the number of switches between thinking activities, time spent on task and the average duration of the thinking activities. The findings from the qualitative study support that the measurements we used for the process variables were valid: where implicit monitoring was seen to lessen in the quantitative study, explicit monitoring also lessened according to the qualitative studies; time on task lessened and the duration of the thinking activities increased as their motivation to persevere dwindled and demonstrations of grit made place for more negativity.

Was the inclusion of the qualitative studies of added value then? We belief so, for now we not only know that there is an effect but we also can explain why the effect occurs. From this knowledge one might progress to developing interventions to improve Latin students' academic buoyancy. A reason to include the qualitative analysis more prominently than the quantitative analysis, is that we were able to closer mimic what teachers see during their lessons and thus make it more relevant to their practice. By qualitatively approaching the data and creating strategy profiles, we assume that teachers are more able to recognise their own students in the descriptions than in a purely quantitative approach. By doing so, we hope we have created a bridge between the practice of educational researchers and teachers. The creating of this bridge between researchers and teachers is one of the main aims of the *Promotiebeurs for Leraren*, the NWO-grant that funded this dissertation.

A counterargument could be made that the inclusion of the quantitative study was unnecessary and that by excluding it there would have been opportunities to delve deeper into the translation processes of all 16 participants. Particularly as we did create process bars and coded the verbal transcripts of all 16 participants. However, the research question included to which extent academic buoyancy was moderated by the learner variables. The quantitative study gave us a focus for what to look out for in the qualitative analyses and which participants to include as the most relevant cases (i.e. the most and least proficient translators and those with a most outspoken mindset).

### *2.4 Individual differences*

At the onset of the study, we included three learner variables to account for individual differences between the participants. We expected these differences to moderate the effect of the setback task on the participants accuracy and translation process. After Chapter 4 we did not further include frustration tolerance. For mindset and translation proficiency we made use of a self-report questionnaire. We realise that this may seem contradictory to our decision and arguments to not use self-report as a measure for academic buoyancy. However, not making use of any self-report questionnaire to measure psychological phenomena is not realistic, as it is time consuming and possibly counterproductive. Self-report is a method that is common in psychology research, in spite of validity concerns. We decided to focus on academic buoyancy via a task-based measure, as that was the central topic of the dissertation, and make use of well-used and prior validated self-report for mindset and frustration tolerance. In the coming sections we re-evaluate how we measured each learner variable.

### 2.4.1 Translation proficiency

A possible critique of the methodology of our study concerns the measurement of the participants' translation proficiency. To determine how proficient the participants were at translating Latin, we made use of their accuracy scores of the pre-setback task. This is slightly tricky, as it meant that the pre-setback task not only functioned as the baseline for the participants' translation processes before the intervention of the setback, but also as a measurement for a learner variable.

Despite this ambiguity, we still assume that this was the best method available to us for measuring the participants translation proficiency. Latin school grades would not suffice, as they also represent other Latin skills and knowledge than translation. Making use of a school grade specifically generated from a translation assignment, might have provided an opportunity. An example of this method can be found in Newland (2016), where previous translation results were used to recognise low proficiency. However, our participants attended different schools, which meant that neither the tests themselves nor the grading would have been equivalent. A third option to determine translation proficiency could have been to create our own translation assignment which all the participants completed and we graded. However, we also decided against this option, as by using part of the Latin Buoyancy Task, our measurement of proficiency was as close as it possibly could be to circumstances and the post-setback task. For, translating whilst thinking-aloud in the presence of an unknown researcher are factors that might influence the quality of the translation. Our study is not alone in the domain of Latin translation that used the pre-test scores to determine proficiency. Luger (2020) included pre-test scores to assess the participants' Latin translation proficiency aiming to establish a baseline.

Thus, we maintain that to measure translation proficiency under circumstances that were particularly comparable to the administering of the Latin Buoyancy Task led to the best validity of the learner variable 'translation proficiency'.

### 2.4.2 Mindset preference

In our study, we included mindset preference as a learner variable. We were mainly interested in the participants with the most outspoken mindset preferences. When we included all the participants' mindset preferences, we approached the scale relatively and not as an absolute scale: Participant A was more inclined to a fixed mindset than Participant B. Our aim was thus not to classify participants as fixed or growth, but to rank the participants as either more or less inclined towards a fixed mindset.

We measured the participants' mindset preferences via a Dutch translation of a three-item questionnaire previously developed by Dweck, Chiu & Hong (1995). This questionnaire has been used in many previous studies (e.g. Blackwell, Trzesniewski & Dweck, 2007; Romero, et al., 2014; Yeager et al., 2019). A recent German study aimed at validating the three-item questionnaire in a German context found that the validity and psychometric properties of the scale were satisfactory (Rammstedt, Grüning & Lechner, 2021). The validity of the questionnaire which formed the source text of our translation, has thus been previously attested.

Our results did differ from Dweck's studies in that our participants proportionally reported more growth mindset preference. Our mean (3.71) was somewhat comparable to the 3.89 ($N$ = 360) found by Rammstedt, Grüning & Lechner (2021). We aimed to check that the items were related to the concept of mindset and gain further understanding of the participants' answers by including questions on the participant's filled-in questionnaire after the retrospective thinking aloud. We specifically asked the participants to elaborate on their answers. An example of this can be found in Quote 7.1. In this example, Horatia explains that intelligence is malleable and dependent on how much effort ('if you do your best') you put in. Her mean score (5.33) agrees with this explanation. On the other hand, Anthony, whose mean score (2.33) corresponded with a fixed mindset, referred to intelligence as 'something you get from your parents at birth and cannot change much[1]'. The explanations the participants gave for their answers agreed with the mindset theory as proposed by Dweck (e.g. Dweck, 1999; Molden & Dweck, 2006; Yeager & Dweck, 2012).

[1] *Iets wat je van je ouders meekrijgt bij je geboorte en dat kan niet echt veranderen.*

> Researcher: You say that you totally do not agree that one has a certain amount of basic intelligence…
> Horatia: You do have a basic amount of intelligence, but I donot think that you cannot do anything about it, because if you just do your best, you automatically understand more. You can be born with the knowledge of vmbo, but just do vwo.

*Quote 7.1. Example of a student's Mindset questionnaire elaboration*[1]

However, Horatia's reference to 'vmbo' and 'vwo' does indicate that cultural bias might have affected our participants' interpretation of the word 'intelligence' and thus the outcome.[2] Cultural differences can affect the generalizability of measurements such as the Dweck questionnaire (Taber, 2014). Other studies that made use of a translated version of a mindset questionnaire also found results that differed from the original. For example, a French study by Cury et al. (2006) found only a -.36 correlation between fixed and growth mindset, instead of -.80 (Hong, et al. 1999). It seems that the Dutch educational system with its different levels of education and the possibility to move 'up' or 'down' affected how participants interpret intelligence. However, the growth mindset participants do refer to effort in relation to intelligence as exemplified by Horatia's quote. On the other hand, the reference to birth is indicative of a fixed view of intelligence. Despite a different overall outcome regarding the proportions between reported growth and fixed mindsets, when asked to put their view in their own words, the participants did speak of intelligence in the same terms of 'malleable' versus 'fixed'.

Moreover, we did not aim to select participants that were an average representation: our participants were purposefully selected via extreme sampling regarding their academic buoyancy. Other factors that connected all our participants was they were gifted and in their third year of Latin. Other studies related to determining the mindset preference of specific gifted participants also found relatively more participants reporting a growth mindset preference (c.f.

---

[1] ***Researcher:*** *Je zegt dat je het helemaal niet er mee eens bent dat je een bepaalde basisintelligentie hebt …* ***Horatia:*** *Je hebt wel een basisintelligentie, maar ik denk niet datje er niets aan kunt doen, want als jij gewoon je best doet ga je vanzelf veel meer snappen zeg maar. Je kunt geboren zijn met de kennis van vmbo, maar gewoon vwo gaan doen.*

[2] *VMBO and VWO are terms related to the Dutch secondary education system. VMBO refers to pre-vocational education, whereas VWO refers to pre-university education. Within Dutch general education, these two types of education are on opposite ends of the spectrum.*

Ziegler & Stoeger, 2018; Mofield & Parker Peters, 2018). Thus, despite finding relatively more participants reporting a growth mindset preference than the instrument when used in its original American setting, we are confident that for our purposes of measuring mindset the validity of the Dutch mindset questionnaire was satisfactory.

### 2.4.3 Frustration tolerance

We tested the Frustration tolerance questionnaire on a relatively small group of participants (*N* = 81). However, the first administration of the questionnaire seemed to provide promising results. The questionnaire itself was an adaptation of a questionnaire by Wright, Lam and Brown (2009). Our results from Chapter 4 indicated that there was a significant correlation between participants who reported having a low frustration tolerance and those who spent significantly less time on task after the setback. According to Wright, Lam and Brown (2009) the motivation of students with a low frustration tolerance decreases when challenged. Results by Farley, Risko and Kingstone (2013) confirmed task disengagement in participants with a low frustration tolerance. Our observed decreased time on task seems to imply that the questionnaire indeed measured frustration tolerance.

Seeing that we assume that the validity of the Frustration tolerance questionnaire was sufficient, one might ask why we did not find an interaction between frustration tolerance and the extent of academic buoyancy in Chapter 4. We think that this is because our participants did experience the setback, but that it did not truly frustrate them. In our quest to make the circumstances as beneficial to thinking aloud as possible, we aimed at making the participants feel as much at ease as possible, for example by emphasizing that the thinking aloud session was not part of a test and no report would be shared with their teacher. We did so, following the procedures for generating reliable data from think aloud, as described by for example Ericsson & Simon (1993), Van Someren, Barnard & Sandberg, (1994) and Young (2005). That Marcus specifically said that he was not stressed can be taken as an indication of that we succeeded in making the participants feel at ease. However, by making them feel at ease, we were also minimizing the load on their frustration tolerance.

The main purpose of our study was not to push their frustration tolerance, but to investigate academic buoyancy. Not being able to solve a problem might evoke the emotion ‘frustration’, but frustration is not the same as frustration tolerance. Frustration tolerance comes into play when the attainment of a goal is impeded. For our participants successfully translating the sentences was probably not a goal they had set themselves. That combined with us aiming to set them at ease, seems to have led to frustration tolerance not being a viable learner variable to include in the study. We did see indications of frustration tolerance and academic buoyancy being related, but if one wanted to scrutinize to what extent they are connected, the setting must be made more frustrating, for example by emphasizing that success is important, and the design must not predominantly aim at collecting valid think aloud data. We thus assume that there was no problem with the questionnaire's validity, but that the concept of frustration tolerance did not actually fit the overall design of our study.

## 3 PRACTICAL IMPLICATIONS & RECOMMENDATIONS

We concluded that the heterogeneous nature of Latin texts necessitates a certain level of academic buoyancy. Apparently, not all students are equally well-equipped to deal with the setbacks that the heterogeneity of Latin texts present students with. Students must thus be actively taught how to not be affected by a difficult passage.

To do so, the student first needs to become aware that heterogeneity is inherent to Latin texts, they need to know that it is okay to sometimes get stuck and that getting stuck is not foreboding for the rest of the task. Encountering problems during translating must thus be somewhat normalized by the teacher. The teacher can do so by sharing own experiences, modelling not only how to approach a difficult passage (meta)cognitively but also affectively. Also, activities that make students visualize that consecutive sentences have differing complexity levels might also help create awareness. Publishers of Latin textbooks, particularly those of the lower years might support this by not only aiming their own created Latin texts at consolidating newly learnt grammatical and semantic features, but also aim to fluctuate the complexity of the sentences.

One of the main alterations our participants made to their translation process, particularly when they were high proficient translators, was to let their relief be dominant and lead to them monitoring less. Making students aware of the possible pitfall of letting emotions rule after encountering a difficult passage might also be helpful. Observational learning methods could cultivate this awareness. If students are able to recognise (unhelpful) strategies in others, they may also become aware of their own (unhelpful) strategies. As we encountered low proficient participants learning from the one-on-one think aloud setting, thinking aloud might also be employed as a learning activity to strengthen students' academic buoyancy.

We found that affective learning strategies seem particularly important in confronting and overcoming setbacks. For teachers this means that they must not only focus on the (meta)cognitive development of their students. This asks of teachers to set aside their presumptions and what they think they are observing. In creating room for affective development, asking the ‘why’ question is important: “why is this student doing X?”. What at first sight seems to be obstructive behaviour (for example, not starting the task), might come from a place of fear, and, therefore, actually be an avoidance strategy. Also, presumptions regarding ‘proficient translators’ might have to be re-evaluated. Is a student who seems a proficient translator based on for example their grades indeed proficient if said student is unable to cope with the heterogeneity of Latin texts?

We finally stress the implication of letting the actual learning process take place in the classroom and not at home through homework tasks. By having the challenging parts of a task take place in the classroom, the teacher is there to observe problems and help. In doing so, the teacher should make use of scaffolding techniques: instead of solving the problem for the student, the student needs to experience solving the problem themselves but with the support of a teacher. If we are able to catch unhelpful affective strategies early in the students’ school career, it might be easier to amend them and therethrough generally improve students’ translation skills.

## 4 FURTHER STUDIES

The end of this dissertation provides five points of departure for further studies. Three of which remain closely linked to Latin learning and instruction, whilst the two others could be placed in a broader research field.

First, a future study might center on participants translating the Latin Buoyancy Task, but with the inclusion of a control group. The control group would also translate the pre- and post-setback task, but instead of the second task containing the setback sentences, it would contain three sentences of similar complexity as the pre- and post-setback task. Such a study might further validate the Latin Buoyancy Task as a tasked based instrument that can be used to further explore academic buoyancy.

The current study included third-year Latin participants and made use of unconnected sentences in the Latin Buoyancy Task. A second possibility for further research lies in including older Latin students as participants. In doing so, the researcher might also include a segment from an authentic Latin text as a measurement instead of unconnected sentences. One could even make use of the nationwide exam results to see whether students were prolongedly affected by a complex segment. Including authentic texts would provide opportunities to also include text comprehension in the study and see whether the heterogeneity of the text also influences text comprehension besides accuracy. By developing a Latin Buoyancy Task for older year Latin students, there are also possibilities for a longitudinal study in which the progression of academic buoyant and non-buoyant Latin students is compared and contrasted. We are aware that not all our participants continued with Latin in their curriculum until the final exams (see also The Participants' Epilogue). A longitudinal study might indicate whether there is a relationship between participants who stop with Latin and to what extent they are academically buoyant.

A third direction would be an intervention study, which aims at improving Latin students' extent of academic buoyancy and therethrough the overall quality of their translations. Such an intervention design could be based on the 5C's following the example of Martin, et al. (2010). These 5C's include confidence, coordination, control, composure and commitment. Interventions could be created for lower- and upper years Latin students. However, we expect that strengthening academic buoyancy in the lower years will be the most beneficial. For, affective strategies are often based on previous experiences and beliefs the students hold (Vermunt, 1996). By strengthening academic buoyancy before the students have too many negative experiencing with translating we of the opinion the largest results may be booked in the long term. Therefore, we would advise an intervention study to focus on the lower years.

When moving away from the Latin learning and instruction field, further research could focus on academic buoyancy in other domains. This would provide the opportunity to learn whether the learning strategy patterns and their moderators in Latin translation are general or subject-specific. By moving away from Latin as the task domain, one might also include gifted participants who attend prevocational education (VMBO) and compare their academic buoyancy to those who attend pre-university education (VWO). This could be of particular interest as not all gifted students leave primary school to follow their secondary education at HAVO or VWO level (c.f. Frumau-van Pinxten, Derksen & Peters, 2021). By comparing the extent of academic buoyancy in gifted participants from different levels it might explain to some extent why already by the age of 12 Dutch students who have been recognised as 'gifted' are significantly underachieving. Finally, there would also be an opportunity to compare and contrast gifted participants with non-gifted participants.

A final opportunity for further research is related to mindset in the Dutch context. Our results from the questionnaire did not align with the findings from Dweck's studies. Our preliminary questioning of why the participants answered as they did, seemed to indicate that cultural bias might explain the discrepancy from Dweck's studies (see Chapter 1). Another possible explanation for the discrepancy could lie in the fact our participants were gifted. Ziegler & Stoeger (2010) and Mofield & Parker Peters (2018) concluded that mindset might work differently in gifted students compared to average students. Further research could aim to ascertain in a Dutch context to what extent cultural bias is at play or/and to which extent giftedness affects the results from the mindset questionnaire.

## 5 THE PARTICIPANTS' EPILOGUE

Our participants were selected because they had demonstrated a large extent or a lack of academic buoyancy in the Anagram Task. In that regard, they were not average students. The Latin Buoyancy Task only included one temporary setback, and we already saw that our participants were affected by the setback, for some, this was a prolonged effect in the post-task. However, the Latin Buoyancy Task was created to mimic the Latin texts that students are asked to translate in most Latin lessons. That implies that for Latin students, the setbacks provided in the texts are not a one-time experience but could become chronic. Of our 16 participants we know that ten of them were non-buoyant. This begs the question, what did our participants' (Latin) educational path look like and did it differ from that of the buoyant participants?

A full study of this question lies outside the scope of the current dissertation. However, we have some data about 11 of our participants and their further educational path, through personal communication with their teachers. By sharing this we hope to emphasize the relevance of academic buoyancy for translating Latin on the one hand and education in general.

Of the 16 participants included in the think-aloud study only seven partook in the final Latin exam after completing six years of Latin. Two of our three most proficient participants dropped Latin before the end of the fourth year. The other did not pass the exam despite sitting it. Three participants sat the exam and scored exceptionally high marks (>9.0). These were 'middle mode' proficient in our study but included in our study as they scored particularly buoyantly on the Anagram Task. Two participants, who were both known as cognitively gifted due to their IQ-score being >129, did not finish high school with a VWO diploma. One of these participants did earn a HAVO diploma and one fully dropped out of school. Both these participants were selected for their lack of academic buoyancy. Finally, one high buoyant participant went on to study Classical languages at university.

## 6 CONCLUDING REFLECTIONS

The motivation for this dissertation on academic buoyancy was found in the Latin classroom. As a Latin teacher, I frequently take a moment to just observe students whilst struggling to translate a text into Dutch. To me, this is a fascinating sight: during their struggle, students display so many different types of strategies to deal with, or explicitly not deal with, problems they encounter. Some might stick their teeth in and fanatically try to solve the puzzle, others might not even start the task, saying it's 'boring' and others continuously ask questions. In time, some will even go as far as wanting to drop the subject because it is 'too difficult'. Yet somehow, it is my job to teach all these students, with all their (learnt) strategies to translate Latin texts into Dutch.

Despite all differences between the individual students and their learning strategies, there is one common denominator that connects my students year after year: translating Latin challenges them. Each year I give my students an assignment in which they are asked to write a reflective essay on translating Latin and specifically their translation process. One of the questions they have to answer is what they have learnt from translating Latin. Over the years, none of my students has answered something related to Latin cognitive knowledge. As Figure 7.1, which includes examples written by students just before they commenced their final exams in 2023, demonstrates, their answers mainly pertain to persevering when challenged. Apparently, even the students among the most academically gifted find translating Latin a challenge and learn other important learning skills such as 'persevering' from translating Latin.

This, combined with the heterogeneity of the source texts, provides Latin teachers with the opportunity to nurture and strengthen academic buoyancy in their students and through this give them a life skill. At the time of writing, the Latin (and Greek) curriculum is being reviewed and revised. This is a great motivation for Latin teachers to speak out what makes Latin different from other subjects and why it is important that schools continue offering Latin as part as their curriculum. In my personal opinion, translating Latin texts not only provides cognitively gifted students with the opportunity for 'reflection on my own world through the unknown'[1], but also for 'reflection on myself as a learner through

[1] *reflectie op het eigenene via het vreemde*

the unknown'[1]. With the completion of this dissertation I hope to give pause for thought and to propose that if we want to see fewer of our gifted students underachieving, instead of concerning ourselves with cognitive knowledge, we teachers must bring our and our students' attention to the whole learning process, including the emotions that go with learning.

[1] *reflectie op mijzelf als leerder via het onbekende*

*Figure. 7.1 Examples of what students claim to have learnt from Latin*[1]

"Translating is extremely difficult for me. I do not have much self- confidence, so I immediately think I can not do it. Whereas, whilst translating you need to persist to solve the puzzle (a sentence). That's why I do not like translating: I am not good at it."

-**Roos Carlier**

"I generally do not enjoy translation as it is super difficult. Through it I have learnt about myself that I am not very good at persisting when I find something difficult and I have learnt how to push through it anyway. This is the most important thing I have learnt, particularly for my further education and life. I do not think I could have learnt this from another subject, because I find Latin difficult, but all you can do is keep going at it."

-**Rosan Honing**

"Latin has really taught me not to give up. Sometimes I got totally stuck and the more it happened, I became all the more frustrated, but in the end you could always work it out."

-**Amy van Ingen**

"Latin has taught me that when I really apply myself to something, I can succeed. In my further educations this will be important as it has given me confidence in my abilities. I do not think that I could have learnt this from another subject, because the serious effort needed to succeed was not necessary in any of my other subjects."

- **Lars van Lieshout**

"Next year I hope to start 'electrical engineering' at the TU Delft. There I am not going to need my Latin knowledge. However, the perseverance and dealing with setbacks will hopefully help me through my university degree. Personally, I do not think I could have learnt this perseverance through another subject. A Latin text is one big puzzle, that you can not simply learn and summarize in the same way as all the other subjects."

- **Wessel Wagemaker**

"Because I find translating Latin quite complex, I am inclined to slacken. I have undergone the consequences of doing so and have learnt that particularly when I find something difficult I have to spend more time on it and give up less quickly."

- **Aukje Wiekmeijer**

[1] For the original Dutch texts, see Appendix 7.1.

## REFERENCES


Abd-El-Fattah, S. M. & Yates, G. C. R. (2006). *Implicit Theory of Intelligence Scale: Testing for factorial invariance and mean structure.* Paper presented at the Australian Association for Research in Education Conference.

Agasisti, T., Avvisati, F., Borgonovi, F. & Longobardi, S. (2018). Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA. *OECD Education Working Papers,* 167, OECD Publishing. https://doi.org/10.1787/e22490ac-en

Al Qahtani, A. (2020). Investigating metacognitive think-aloud strategy in improving Saudi EFL learners' reading comprehension and attitudes. *English language teaching, 13*(9), 50-62. https://doi.org/10.5539/elt.v13n9p50

Albaili, M. A. (2003). Motivational goal orientations of intellectually gifted achieving and underachieving students in the United Arab Emirates. *Social Behavior and Personality: An International Journal, 31*(2), 107-120. https://doi.org/10.2224/sbp.2003.31.2.17

Alexopoulou, A., Batsou, A. & Drigas, A. (2019). Resilience and Academic Underachievement in Gifted Students: Causes, Consequences and Strategic Methods of Prevention and Intervention. *International Journal of Online and Biomedical Engineering*, *15*(14), 78-86. https://doi.org/10.3991/ijoe.v15i14.11251

American Psychological Association. (2015). *APA Dictionary of psychology.* (G. R. van den Bos, Ed.; 2nd Edition.). Maple Press.

Anderson, C. A. & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53,* 27-51. http://dx.doi.org/10.1146/annuref.psych.53.100901.135231

Arnau, S., Brümmer T., Liegel, N., & Wascher, E. (2021). Inverse effects of time-on-task in task-related and task-unrelated theta activity. *Psychophysiology*, *58*(6). https://doi.org/10.1111/psyp.13805

Aron, A. & Aron, E. (1999). *Statistics for Psychology* (2nd ed.). Prentice Hall.

Aslam, S, & Ali, M. S. (2017). Effect of Self-Efficacy on Students' Achievement in Science: A Case of Secondary School Students in Pakistan. European Journal of Education Studies, 3, 220-235. htttps://doi.org/10.5281/zenodo.1050292

Aspinwall, L. G. & Richter, L. (1999). Optimism and self-mastery predict more rapid disengagement from unsolvable tasks in the presence of alternatives. *Motivation and Emotion, 23*(3), 221-245. https://doi.org/10.1023/A:1021367331817

Austin, J. & Delaney, P.F. (1998). Protocol Analysis as a Tool for Behavior Analysis. *Analysis for Verbal Behavior, 15*, 41–56. https://doi.org/10.1007/BF03392922

Bala, I., Kaur, R. & Singh, S. (2017). Self-efficacy of senior secondary school students with respect to demographic variables. *International Journal of Advanced Research and Development, 2*, 111-114. https://doi.org/10.33394/jp.v10i1.5798

Balduf, M. (2009). Underachievement among college students. *Journal of Advanced Academics, 20*, 274-294. https://doi.org/10.1177/1932202X0902000204

Bannert, M. & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning, 3*(1), 39–58. https://doi.org/10.1007/s11409-007-9009-6

Barbier, K., Donche, V. & Verschueren, K. (2019). Academic (under)achievement of intellectually gifted students in the transition between primary and secondary education: An individual learner perspective. *Frontiers in Psychology, 10*, 1-12. https://doi.org/10.3389/fpsyg.2019.02533

Barnett A. G., van der Pols, J. C. & Dobson, A. J. (2004). Regression to the mean: what it is and how to deal with it, *International Journal of Epidemiology*, *34(1),* 215–220. https://doi.org/10.1093/ije/dyh299

Barrett, L. F., Lindquist, K. A. & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences, 11*(8), 327-32. https://doi.org/10.1016/j.tics.2007.06.003.

Bartelds, D. (2021). How To Stay in the Loop. A Think-Aloud Study on Dictionary Use by Excellent Secondary-School Students of Ancient Greek, *International Journal of Lexicography*, *34*(4), 453–471. https://doi.org/10.1093/ijl/ecab001

Baslanti, U. & McCoach, D. B. (2006). Factors related to the underachievement of university students in Turkey. *Roeper Review, 28*, 210-215. https://doi.org/10.1080/02783190609554366

Baumanns, L. & Rott, B. (2022). The process of problem posing: Development of a descriptive phase model of problem posing. *Educational Studies in Mathematics*, *110*, 251–269. https://doi.org/10.1007/s10649-021-10136-y

Baumeister, R. F. & Tierney, J. (2012). *Die Macht der Disziplin: Wie wir unseren Willen trainieren können* [The power of discipline: How we can train our will]. Campus Verlag.

Bazeley, P. (2013). *Qualitative Data Analysis with NVivo*. SAGE Publications.

Bennett-Rappell, H. & Northcote, M. (2016). Underachieving Gifted Students: Two Case Studies. *Issues in Educational Research, 26*, 407-430.

Berkowitz, L. (1989). Frustration-Aggression Hypothesis: Examination and Reformulation. *Psychological Bulletin, 106*, 59-73. https://doi.org./10.1037/0033-2909.106.1.5

Betts, G.T. & Neihart, M. (1988). Profiles of the gifted and talented. *Gifted child quarterly, 32*(2), 248-253. https://doi.org/10.1177/001698628803200202

Betts, G.T. & Neihart, M. (2010). Revised profiles of the gifted and talented. Retrieved from: http://ingeniosus.net/wp-content/uploads/2010/11/PROFILES-BEST-REVISED-MATRIX-2010.pdf

Beyer, B. K. (2000). *Improving Student Thinking: A Comprehensive Approach.* Allyn & Bacon.

Białek, M. & Domurat, A. (2017). Cognitive abilities, analytic cognitive style and overconfidence: a commentary on Duttle (2016). *Bulletin of Economic Research. 70*. https://doi.org/10.1111/boer.12117.

Bird, C. M. (2005). How I Stopped Dreading and Learned to Love Transcription. *Qualitative Inquiry*, *11* (2), 226–248. https://doi.org/10.1177/1077800404273413

Blackwell, L. A., Trzesniewski, K. H. & Dweck, C. S. (2007). Theories of intelligence and achievement across the junior high school transition: A longitudinal study and an intervention. *Child Development, 78*, 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Boaler, J. & Dweck, C. S. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages, and innovative teaching.* Jossey-Bass.

Boekaerts, M. & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*, 417–450. Academic Press. https://doi.org/10.1016/B978-012109890-2/50042-1

Bokhove, C. & Downey, C. (2018). Automated generation of "good enough" transcripts as a first step to transcription of audio-recorded data. *Methodological Innovations, 11*, 1-14. https://doi.org/10.1177/2059799118790743.

Bouman, D. E. (2011). Frustration Tolerance. In: J. S. Kreutzer, J. De Luca & B. Caplan (Eds.) *Encyclopedia of Clinical Neuropsychology.* Springer, New York, NY. https://doi.org/10.1007/978-0-387-79948-3_2139

Bowles, M.A. (2010). The Think-Aloud Controversy in Second Language Research (1st ed.). Routledge. https://doi.org/10.4324/9780203856338

Boyd, R. M. (2018). Latin students' bottom-up and top-down strategies for reading Latin literature and the impact of cross-linguistic influence. *Journal of Latin Linguistics, 17*(2), 301-332. https://doi.org/10.1515/joll-2018-0014

Branch, J. L. (2000). The trouble with Think Alouds: Generating Data Using Concurrent Verbal Protocols. *CAIS 2000: Dimensions of a Global Information Science.* Canadian Association for Information Science Proceedings of the 28th Annual Conference.

Bridges, K. R. & Roig, M. (1997). Academic procrastination and irrational thinking: A re-examination with context controlled. *Personality and Individual Differences, 22*(6), 941-944. https://doi.org/10/1016/S0191-8869(96)00273-5

Brinkman, J. A. (1993). Verbal protocol accuracy in fault diagnosis. *Ergonomics, 36*(11), 1381–1397. https://doi.org/10.1080/00140139308968007

Brislin, R. W. (1986). The wording and translation of research instruments. In: W. J. Lonner & J. W. Barry (Eds.), *Field methods in cross-cultural research, 7*, 137-164, SAGE Publications. https://doi.org/10.2307/2349027.

Brown, R. & Pressley, M. (1994). Self-regulated reading and getting meaning from text: The Transactional Strategies Instruction model and its ongoing validation. In D. H. Schunk and B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications*, 155–180. Erlbaum.

Bucholtz, M. (2000). The politics of transcription. *Journal of pragmatics, 32*, 1439-1465. https://doi.org/10.1016/S0378-2166(99)00094-6

Buijs, A. (2017). *Statistiek om mee te werken* [Statistics to work with]. Noordhoff.

Byrka, K., Cantarero, K., Dolinski, D. & Van Tilburg, W. (2021). *Consequences of Sisyphean Efforts: Meaningless Effort decreases motivation to engage in subsequent conservation behaviors through disappointment. Sustainability, 13*(10), 1-27. https://doi.org/10.3390/su13105716

Carr, M. & Taasoobshirazi, G. (2008). Metacognition in the gifted: Connections to expertise. In M. F. Shaugh- nessy, M. V. J. Veenman, & C. Kleyn-Kennedy (Eds.), *Meta-Cognition: A recent review of research, theory and perspectives,* 109-125. Nova Science Publishers.

Carr, M., Borkowski, J. G. & Maxell, S. E. (1991). Motivational components of underachievement. *Developmental Psychology, 27*(1), 108-118. https://doi.org/10.1037/0012-1649.27.1.108

Celis Rangel, J. G., King, M. & Muldner, K. (2020). An incremental mindset intervention increases effort during programming activities but not performance. *ACM Transactions on Computing Education, 20*(2), 1-18. https://doi.org/10.1145/3377427

Charness, G., Gneezy, U. & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009g

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal, 12*(2), 68-82. https://doi.org/10.26522/brocked.v12i2.38

Christie, D. (2005). Introduction to IQ testing. *Psychiatry, 4*(6), 22-25 https://doi.org/10.1383/psyt.4.6.22.66351

Clark, L. A., Cuthbert, B,, Lewis-Fernández, R., Narrow, W. E. & Reed, G. M. (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest, 18* (2), 72-145. https://doi.org/10.1177/1529100617727266.

Cohors-Fresenborg, E. & Kaune, C. (2007). Modelling classroom discussions and categorizing discursive and metacognitive activities. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education*, 1180–1189. ERME.

Comerford, J., Batteson, T. & Tormey, R. (2015). Academic buoyancy in second level schools: Insights from Ireland. *Social and Behavioral Sciences, 197*(1) 98-103. https://doi.org./10.1016/j.sbspro.2015.07.061

Cook, D. & Artino, A. (2016). Motivation to learn: an overview of contemporary theories. *Medical Education,* (50), 997-1014. https://doi.org/10.1111/medu.13074.

Coughlan, T., Lister, K. & Lucassen, M. (2021) Representing the Unseen with "Our Journey": A Platform to Capture Affective Experiences and Support Emotional Awareness in University-Level

Study. *Journal of Formative Design in Learning 5*, 39–52. https://doi.org/10.1007/s41686-021-00055-9

Craig, K., Hale, D., Grainger, C. & Stewart, M. E. (2020). Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition Learning 15*, 155–213. https://doi.org/10.1007/s11409-020-09222-y

Cramond, B. (2004). Can We, Should We, Need We Agree on a Definition of Giftedness? *Roeper Review: A Journal on Gifted Education, 27*(1), 15–16. https://doi.org/10.1080/02783190409554282

Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative approaches to research*. (2nd ed.) Merrill/Pearson Education.

Creswell, J. W. (2009). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (3rd ed.). SAGE Publications.

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.). SAGE Publications.

Creswell, J. W. & Plano Clark, V.L. (2011). *Designing and conducting mixed method research.* (2nd ed.). SAGE Publications.

Cromley, J. G. & Kunze, A. J. (2020). Metacognition in education: Translational research. *Translational Issues in Psychological Science, 6*(1), 15–20. https://doi.org/10.1037/tps0000218

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. Jossey-Bass Publishers.

Cury, F., Da Fonseca, D., Elliot, A. J. & Moller, A. C. (2006). The social-cognitive model of achievement motivations and the 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology,* 9(4), 666-679. https://doi.org/10.1037/0022-3514.90.4.666

D'Mello, S. K. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105,* 1082-1099. http://dx.doi.org/10.1037/a0032674

Da Silva, J. (2021). Producing 'good enough' automated transcripts securely: Extending Bokhove and Downey (2018) to address security concerns. *Methodological Innovations, 14,* 1-11. https://doi.org/205979912098776. 10.1177/2059799120987766.

Dai, D. Y. (2000). To be or not to be (challenged), that is the question: Task and ego orientations among high-ability, high achieving adolescents. *The Journal of Experimental Education, 68*, 311–330. https://doi.org/10.1080/00220970009600641

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*, 105–134. https://doi.org/10.1007/s11251-009-9110-0

Dek, J. E. & Kooij, A. P. (2012). *De RAKIT-2: Veranderingen ten opzichte van de oorspronkelijke RAKIT* [The RAKIT-2: Changes compared to the original RAKIT]. Pearson.

Diezmann, C. M. & Watters, J. J. (2002). The importance of challenging tasks for mathematically gifted students. *Gifted and Talented International, 17*(2), 76-84. https://doi.org/10.1080/15332276.2002.11672991

Duckworth, A. L. (2009). (Over and) beyond high-stakes testing. *American Psychologist, 64*(4), 279-280. https://doi.org/10.1037/a0014923

Duckworth, A.L, & Gross, J. J. (2014). Self-control and grit: related but separable determinants of success. *Current Directions in Psychological Science, 23*(5), 319-325. https://dx.doi.org/10.1177/0963721414541462

Duckworth, A. L., Peterson, C., Matthews, M. D. & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087-1101.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment, 91*(2), 166-174. https://doi.org/10.1080/00223890802634290

Duckworth, A. L. & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44,* 237-251. https://doi.org/10.3102/0013189X15584327

Dupeyrat, C. & Mariné, C. (2015). Implicit theories of intelligence, goal orientation, cognitive engagement, and achievement: A test of Dweck's model with returning to school adults. *Contemporary Educational Psychology, 30*(1), 43-59. https://doi.org/10.1016/j.cedpsych.2004.01.007

Dweck, C. S. (1999). *Self-Theories: Their role in motivation, personality and development.* Taylor & Francis.

Dweck, C. S. (2011). Implicit Theories. In P. Van Lange, A. Kruglanski & T. Higgins (Eds.). *The Handbook of Theories of Social Psychology*. (43-61). SAGE Publications.

Dweck, C. S. (2012). Mindsets and malleable minds: Implications for giftedness and Talent. In R. F. Subotnik, A. Robinson, C. M. Callahan & E. J. Gubbins (Eds), *Malleable minds: Translating insights from psychology and neuroscience to gifted education,* (7-18). Storrs.

Dweck, C. S., Chiu C. Y., & Hong, Y. Y. (1995). Implicit Theories: Elaboration and extension of the model. *Psychological inquiry, 6* (4), 322-333. https://doi.org/10.1207/s15327965pli0604_12

Dweck, C. S. & Legget, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256-273. https://doi.org/10.1037/0037/0033-295X.95.2.256

Efklides, A., Papadaki, M., Papantoniou, G. & Kiosseoglou, G. (1997). Effects of Cognitive Ability and Affect on School Mathematics Performance and Feelings of Difficulty. *The American journal of psychology, 110*, 225-58. http://dx.doi.org/10.2307/1423716.

Efklides, A. & Kourkoulou, A. & Mitsiou, F. & Ziliaskopoulou, D. (2006). Metacognitive knowledge of effort, personality factors, and mood state: Their relationships with effort-related metacognitive experiences. *Metacognition and Learning, 1*, 33-49. https://doi.org/10.1007/s11409-006-6581-0.

Efklides, A., Schwartz, B. L. & Brown, V. (2018). Motivation and affect in self-regulated learning: Does metacognition play a role? In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance,* (64–82). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-5

Eikeboom, R. (1970). *Rationales Lateinlernen*. Vandenhoeck & Ruprecht.

Elliot, A. J. & McGregor, H. A. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 503-519. https://doi.org/10.1037/0022-3514.80.3.501

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Ericsson, K. A. & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*, Revised Edition. The MIT Press.

Eyre, D. (1997). *Able children in ordinary schools.* David Fulton.

Fan, X., Miller, B. C., Park, K., Winward, B. W., Christensen, M., Grotevant, H. D. & Tai, R. H. (2006). An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods, 18*(3), 223-244. https://doi.org/10.1177/152822X06289161

Faria, L. & Fontaine, A. M. (1997). Adolescents' personal conceptions of intelligence: The development of a new scale and some exploratory evidence. *European Journal of Psychology of Education, 12*, 51-62. https://doi.org/10.1007/BF03172869

Farley, J., Risko, E. F. & Kingstone, A. (2013). Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering. *Frontiers in Psychology, 18*(4), 1-9. https://doi.org/10.3389/fpsyg.2013.00619

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement, 5*(1), 105–112. https://doi.org/10.1177/014662168100500115

Florian, L. (2015). *Heimliche Strategien: Wie übersetzen Schülerinnen und Schüler?* [Secret strategies: How do students translate?]. V&R Unipress.

Fornia, G. L. & Frame, M. W. (2001). The Social and Emotional Needs of Gifted Children: Implications for Family Counseling. *The Family Journal*, *9*(4), 384–390. https://doi.org/10.1177/1066480701094005

Fox, M. C., Ericsson, K. A. & Best, R. (2011). Do Procedures for Verbal Reporting of Thinking Have to Be Reactive? A Meta-Analysis and Recommendations for Best Reporting Methods. *Psychological Bulletin, 137*(2), 316-344. https://doi.org/10.1037/a0021663

Fried, L. & Chapman, E. (2012). An investigation into the capacity of student motivation and emotion regulation strategies to predict engagement and resilience in the middle school classroom. *The Australian Educational Researcher, 39*, 295-311. https://doi.org/10.1007/s13384-011-0049-1

Frumau-van Pinxten, M., Derksen, J. & Peters, W. (2021). Openness to Experience, a Personality Trait of Gifted Adolescents and One of the Key Factors of High Developmental Potential. *International Journal of Secondary Education*, *9*(3), 74-85. https://doi.org/10.11648/j.ijsedu.20210903.11

Futch, O. (1935). A study of eye-movements in the reading of Latin. *The Journal of General Psychology, 13*(2), 434-463. http://dx.doi.org/10.1080/00221309.1935.9917896

Gagné, F. Y. (2000). Transforming gifts into talents: The DMGT as a developmental theory 1. *High Ability Studies, 15*, 119-147. http://dx.doi.org/10.1080/1359813042000314682

Gagné, F. Y. (2003). Transforming gifts into talents: The DMGT as a developmental theory. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed.), 60-74. Allyn & Bacon.

Gerhards, L. & Gravert, C. (2020). Because of you I did not give up – Peer effects in perseverance. *Journal of Economic Psychology, 81*, 1-33. http://dx.doi.org/10.1016/j.joep.2020.102316.

Ghisi, M., Bottesi, G., Re, A. M., Cerea, S. & Mammarella, I. C. (2016). Socioemotional Features and Resilience in Italian University Students with and without Dyslexia. *Frontiers in Psychology*, *7*, 478-491. https://doi.org/10.3389/fpsyg.2016.00478

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*(3), 608–626. https://doi.org/10.1037/a0034716

Goossens, M. (2004). *Studiebegeleiding en communicatie bij hoogbegaafde leerlingen* [Study guidance and communication for gifted students]. In T. Kieboom & A. Hermans (Eds.) Hoogbegaafde leerlingen op de secundaire school: hoogvliegers of kwetsbare vogels?. 91-101. Garant.

Göpferich, S., Jakobsen, A. L. & Mees, I. M. (2009). *Behind the mind: Methods, models and results in translation process research*. Samfundslitteratur.

Green, J., Franquiz, M. & Dixon, C. (1997). The Myth of the Objective Transcript: Transcribing as a Situated Act. *TESOL Quarterly*, *31* (1), 172–176. https://doi.org/10.2307/3587984

Greene, J. A., Hutchison, L. A., Costa, L. J. & Crompton, H. (2012). Investigating how college students' task definitions and plans relate to self-regulated learning processing and understanding of a complex science topic. *Contemporary Educational Psychology*, *37*, 307–320.

Greene, J. A., Moos, D. C., Azevedo, R. & Winters, F. I. (2008). Exploring differences between gifted and grade-level students' use of self-regulatory learning processes with hypermedia. *Computers & Education, 50*(3), 1069-1083. https://doi.org/10.1016/j.compedu.2006.10.004.

Grobman, J. (2006). Underachievement in gifted adolescents and young adults: A psychiatrist's view. *Journal of Secondary Gifted Education, 17*(4), 199-210. https://doi.org/10.4219/jsge-2006-408

Gross, M. U. M. (1993). *Exceptionally Gifted Children. Routledge.*

Guan, Z., Lee, S., Cuddihy, E. & Ramey, J., (2006). The validity of the stimulated retrospective think-aloud method as measured by eye-tracking. *Proceedings of the SIGCHI conference on human factors in computing systems*, 1253-1263. https://doi.org/10.1145/1124772.1124961

Güss, C. D. (2018). What is going through your mind? Thinking aloud as a method of cross-cultural psychology. *Frontiers in Psychology, 9,* 1293. https://doi.org/10.3389/fpsyg.2018.01292

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2006). *Multivariate data analysis*. Prentice Hall Pearson Education.

Harkness, J. A., Villar, A. & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts,* John Wiley & Sons, 117–140. https://doi.org/10.1002/9780470609927.ch7

Harrington, N. (2005). It's too difficult! Frustration intolerance beliefs and procrastination. *Personality and Individual Differences, 39*(5), 873-883. https://doi.org/10.1016/j.paid.2004.12.018

Heirweg, S., De Smul, M., Devos, G. & Van Keer, H. (2019). Profiling upper primary school students' self-regulated learning through self-report questionnaires and think-aloud protocol analysis. *Learning and Individual Differences, 70,* 155–168. https://doi.org/10.1016/j.lindif.2019.02.001

Hennessy, M., Dennehy, R., Doherty, J. & O'Donoghue, K. (2022). Outsourcing Transcription: Extending Ethical Considerations in Qualitative Research. *Qualitative Health Research, 32*, 1197-1204. https://doi.org/10.1177/10497323221101709.

Herman, B. C. (2015). The influence of global warming science views and sociocultural factors on willingness to mitigate global warming. *Science Education, 99*(1), 1-38. https://doi.org/10.1002/sce.21136

Hom, H. L. & Maxwell, F. R. (1983) The impact of task difficulty expectations on intrinsic motivation. *Motivation and Emotion, 7*(1), 19-24. https://doi.org/10.1007/BF00992962

Honeyfield, J. (1993). Responding to task difficulty. In M. Tickeroo (ed.), *Simplification: Theory and Practice.* 127-138. Singapore: Regional Language Center.

Hong, Y., Chiu, C. Y., Dweck, C. S., Lin, D. & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*, 588-599. https://doi.org/10.1037/0022-3514.77.3.588

Hoza, B., Waschbusch, D. A., Owens, J. S., Pelham, W. E. & Kipp, H. (2001). Academic task persistence of normally achieving ADHD and control boys: Self-evaluations, and attributions. *Journal of Consulting and Clinical Psychology, 69*(2), 271–283. https://doi.org/10.1037/0022-006X.69.2.271

Jahedizadeh, S., Ghonsooly, B. & Ghanizadeh, A. (2019). Academic buoyancy in higher education: Developing sustainability in language learning through encouraging buoyant EFL students. *Journal of Applied Research in Higher Education, 11*(2), 162-177. https://doi.org/10.1108/JARHE-0402918-0067

Jonassen, D. H. & Grabowski, B. L. (2012). *Handbook of individual differences learning and instruction.* Routledge.

Karten, S. (2015). *Ovidius lezen: voor iedereen net even anders. Een onderzoek naar het effect van differentiëren op de leesvaardigheid en motivatie van leerlingen in de bovenbouw* [Reading Ovid: Just a bit different for everyone: A study on the effect of differentiation of reading skills and student motivation in the upper grades]. *Lampas*, *48*(3), 314–325.

Kemper, E., Stringfield, S. & Tetley, C. (2003). Mixed methods sampling strategies in social science research. In A. Tashakkori and C. Teddlie (Eds.), *Handbook of mixed methods in social & behaviorial research.* (273-296). SAGE Publications.

Kesler, T., Tinio, P.P., & Nolan, B.T. (2016). What's Our Position? A Critical Media Literacy Study of Popular Culture Websites with Eighth-Grade Special Education Students. *Reading & Writing Quarterly, 32*, 1-26. https://doi.org/10.1080/10573569.2013.857976

Keys, C. W. (2000). Investigating the thinking processes of eighth grade writers during the composition of a scientific laboratory report. *Journal of Research in Science Teaching*, *37*(7), 676-690. https://doi.org/10.1002/1098-2736(200009)37:7<676::AID-TEA4>3.0.CO

Kieboom, T. (2015). *Hoogbegaafd: Als je kind (g)een Einstein is* [Gifted: When your child is (not) Einstein]. Lannoo.

Kim, H. R. & Bowles, M.A. (2019). How deeply do second language learners process written corrective feedback? Insights gained from think- alouds. *TESOL Quarterly*. https://doi.org/10.1002/tesq.522

Kim, H. & Han, K. (2014) Development of academic resilience scale for gifted youth. *Journal of Gifted/Talented Education, 24*(2), 289-312. https://doi.org/10.9722/JGTE.2014.24.2.289

King, R. B. (2017). A fixed mindset leads to negative affect: The relations between implicit theories of intelligence and subjective well-being. *Zeitschrift für Psychologie, 225*(2), 137–145. https://doi.org/10.1027/2151-2604/a000290

Kitchell, K.F. (2000). "Latin III's Dirty Little Secret: Why Johnny Can't Read." *New England Classical Journal* 27: 205–226. https://doi.org/10.52284/NECJ/49.1

Klein, E., Bieck, S. M., Bloechle, J., Huber, S., Bahnmueller, J., Willmes, K. & Moeller, K. (2019). Anticipation of difficult tasks: neural correlates of negative emotions and emotion regulation. *Behavioral and brain functions: BBF*, *15*(1), 4. https://doi.org/10.1186/s12993-019-0155-1

Koenen, M. (2007). *Inleiding tot de Latijnse syntaxis: Structuur van zin en tekst oefenboek* [Introduction to Latin syntax: Sentence and tekst structure workbook]. Amsterdam University Press

Kroon, C. H. M. (2007). *Inleiding tot de Latijnse syntaxis: Structuur van zin en tekst, grammatica* [Introduction to Latin syntax: Sentence and text structure]. Amsterdam University Press.

Kroon, C. H. M. & Sluiter, I. (2010). *Het Geheim van de Blauwe Broer. Eindrapport van de Verkenningscommissie Klassieke Talen* [The secret of the blue brother. Final report by the classical languages exploratory committee]. Stichting Leerplanontwikkeling. http://www.slo.nl/voortgezet/tweedefase/nieuws/rappkt/

Kuhlmann, P. (2015). Lateinische Texte richtig übersetzen – (k)ein Problem? Die lernpsychologischen Voraussetzungen für das Verstehen von lateinischen Texten [Translating Latin texts correctly - (not) a problem? The psychological learning requirements for understanding Latin texts]. In M. Frisch (ed.), *Alte Sprachen – Neuer Unterricht* (Ars Didactica: Marburger Beiträge zu Studium und Didaktik der alten Sprachen 1). 11–34. Kartoffeldruck-Verlag.

Kujamäki, M. (2015). On the operationalisation of 'pauses' in translation process research. *Translation and Interpreting, 7,* 47-58. https://doi.org/10.12807/ti.106201.2015.a04

Kuusela, H. & Paul, P. (2000). A Comparison of Concurrent and Retrospective Verbal Protocol Analysis. *American Journal of Psychology, 113*, 387-404. http://dx.doi.org/10.2307/1423365

Lang, G. & van der Molen, H. T. (1998). *Psychologische gespreksvoering* [Psychological counselling]. Uitgeverij H. Nelissen.

Lapadat, J. (2000). Problematizing transcription: Purpose, paradigm and quality. *International Journal of Social Research Methodology, 3*, 203-219. https://doi.org/10.1080/13645570050083698.

Lawanto, O., Minichiello, A., Uziak, J. & Febrian, A. (2019). Task affect and task understanding in engineering problem solving. *Journal of Technology Education*, *30*(2), 21–38. https://doi.org/10.21061/jte.v30i2.a.2

Leighton, J. P. (2017). *Using think-aloud interviews and cognitive labs in educational research.* Oxford University Press.

Leighton, J. P. (2021). Rethinking Think-Alouds: The Often-Problematic Collection of Response Process Data. *Applied Measurement in Education, 34*(1), 61-74. https://doi.org/10.1080/08957347.2020.1835911

Lens, W. & Rand, P. (2000). Motivation and cognition: Their role in the development of giftedness. In: K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.) *International Handbook of Giftedness and Talent,* (193-202). Elsevier Science.

Leow, R. P. & Morgan-Short, K. (2004). To Think Aloud or Not to Think Aloud: The Issue of Reactivity in SLA Research Methodology. *Studies in Second Language Acquisition, 26*(1), 35-57. https://doi.org/10.1017/S0272263104261022

Lockl, K., & Schneider, W. (2003). Metakognitive Überwachung- und Kontrollprozesse bei der Lernzeiteinteilung von Kindern [Metacognitive monitoring and control processes in children's time management for learning]. *Zeitschrift Für Pädagogische Psychologie*, 17, 173–183. https://doi.org/1010-0652.17.34.173.

Lombard, M., Snyder-Duch, J. & Bracken, C. C. (Eds.). (2017). *The SAGE Encyclopedia of Communications Research Methods* (Vols. 1-4). SAGE Publications.

Lucas, G. M., Gratch, J., Cheng, L. & Marsella, S. (2015). When the going gets tough: Grit predicts costly perseverance. *Journal of Research in Personality, 59,* 15-22. https://doi.org/10.1016/j.jpr.2015.08.004

Luger, S. (2018). How do Dutch adolescents translate Latin into coherent Dutch? A journey into the unknown. *Journal of Latin Linguistics*, *17*(2), 333-365. https://doi.org/10.1515/joll-2018-0015

Luger, S. (2020). Lost in Latin translation. Teaching students to produce coherent target texts. Retrieved from https://dare.uva.nl/search?identifier=159400db-acaa-41a6-8ad1-5a3bc62b6bbe

Lui, J. (2010). Language learning strategies and its training model. *International Education Studies, 3*(3), 100-104.

MacCann, C., Jiang, Y., Brown, L. E., Double, K. S., Bucich, M., & Minbashian, A. (2020). Emotional intelligence predicts academic performance: A meta-analysis. *Psychological Bulletin, 146*(2), 150–186. https://doi.org/10.1037/bul0000219

Macefield, R. (2009). How to Specify the Participant Group Size for Usability Studies: A Practitioner's Guide. *Journal of Usability Studies, 5*(1), 34-45.

Mackenzie, C., Macdougall, C., Fane, J. & Gibbs, L. (2018). Using Emoji in Research with Children and Young People: Because We Can? Vol. 1: Proceedings of the 3rd World Conference on Qualitative Research.

Maddi, S. R., Matthews, M. D., Kelly, D., Villarreal, B. J. & White, M. (2012). The role of hardiness and grit in predicting performance and retention of USMA cadets*. Military Psychology, 24*(1), 19-28. http://dx.doi.org/10.1080/08995605.2012.639672

Mahon, N. E., Yarcheski, A., Yarcheski, T. J. & Hanks, M. M. (2007). Mediational models of health practices in early adolescents. *Clinical Nursing Research, 16*(4), 302–316. https://doi.org/10.1177/1054773807307314

Makel, M. C., Kell, H. J., Lubinski, D., Putallaz, M. & Benbow, C. P. (2016). When Lightning Strikes Twice: Profoundly Gifted, Profoundly Accomplished. *Psychological Science, 27*(7) 1004-1018. https://doi.org/10.1177/0956797616644735

Malek, H. B., Berna, F. & D'Argembeau, A. (2017). Reconstructing the times of past and future personal events. *Memory, 25*, 1402-1411. https://doi.org/10.1080/09658211.2017.1310251

Malmberg, L. E., Hall, J. & Martin, A. J. (2013). Academic buoyancy in secondary school: Exploring patterns of convergence in English, mathematics, science and physical education. *Learning and Individual Differences, 23*(1), 262-266. https://doi.org/10.1016/j.lindif.2012.07014

Malmberg, J., Haataja, E. & Järvelä, S. (2022). Exploring the connection between task difficulty, task perceptions, physiological arousal and learning outcomes in collaborative learning situations. *Metacognition Learning*, *17*, 793–811. https://doi.org/10.1007/s11409-022-09320-z

Martin, A. J. (2013). Academic buoyancy and academic resilience: Exploring 'everyday' and 'classic' resilience in the face of academic adversity. *School Psychology International, 34*(5), 488-500. https://doi.org/10.1177/0143034312472759

Martin, A. J., Colmar, S., Davey, L. & Marsh, H. (2010). Longitudinal modelling of academic buoyancy and motivation: Do the 5Cs hold up over time?. *The British Journal of Educational Psychology*, *80,* 473-96. https://doi.org/10.1348/000709910X486376

Martin, A. J. & Marsh, H. W. (2006). Academic resilience and its psychological and educational correlates: A construct validity approach. *Psychology in the Schools, 43*(3), 267–281. https://doi.org/10.1002/pits.20149

Martin, A. J. & Marsh, H. (2008). Academic buoyancy: Towards an understanding of students' everyday academic resilience. *Journal of School Psychology, 46*(1), 53-83. https://doi.org/10.1016/j.jsp.2007.01.002

Martin, A. J. & Marsh, H.W. (2009). Academic resilience and academic buoyancy: Multidimensional and hierarchical conceptual framing of causes, correlates, and cognate constructs. *Oxford Review of Education*, *35*, 353-370. https://doi.org/10.1080/03054980902934639

Matthews, M. S. & McBee, M. T. (2007). School factors and the underachievement of gifted students in a talent search summer program. *Gifted Child Quarterly, 51*, 167-181. http://dx.doi.org/10.1177/0016986207299473

McBurney, D. H. & White, T. L. (2004). *Research methods* (6th ed.). Nelson.

McCoach, D. B. & Flake, J. K., (2018). The role of motivation. In: S. I Pfeiffer, E. Shaunessy-Dedrick & M. Foley Nicpon (Eds), *APA Handbook of Giftedness and Talent,* (201-213). American Psychological Association. https://doi.org/10.1037/0000038-013

McDonald, J.D. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire, 1*(1), 75-94.

McMullin, C. (2023). Transcription and qualitative methods: Implications for third sector research. *Voluntas, 34,* 140-135. https://doi.org/10.1007/211266-021-00400-3

Meijer, J., Veenman, M. V. J. & van Hout-Wolters, B. H. A. M. (2006). Metacognitive Activities in Text-Studying and Problem-Solving: Development of a Taxonomy. *Educational Research and Evaluation, 12*(3), 209–237. https://doi.org/10.1080/13803610500479991

Meindle, P., Yu, A., Galla, B. M., Quirk, A., Haeck, C., Goyer, J. P., Lejeuz, C. W., D'Mello, S. K. & Duckworth, A. L (2019). A brief behavioral measure of frustration tolerance predicts academic achievement immediately and two years later. Emotion, 19(6), 1081-1092. 10.1037/emo0000492

Meissener, W. W. (2008). The role of language in the development of the self III: The significance of pronouns. *Psychoanalytic Psychology, 25*(2), 242-256. 10.1037/0736-9735.25.2.242

Meneghel, I., Martínez, I. M., Salanova, M. & de Witte, H. (2019). Promoting academic satisfaction and performance: Building academic resilience through coping strategies. *Psychology in the Schools*, *56*, 875–890. https://doi.org/10.1002/pits.22253

Mofield, E. & Parker Peters, M. (2018). Mindset misconception? Comparing mindsets, perfectionism, and attitudes of achievement in gifted, advanced and typical students. *Gifted Child Quarterly, 62*(4), 327-349. https://doi.org/10.1177/0016986218758440

Molden, D. C. & Dweck, C. S. (2006). Finding "Meaning" in Psychology: A Lay Theories Approach to Self-Regulation, Social Perception, and Social Development. *American Psychologist, 61*(3), 192–203. https://doi.org/10.1037/0003-066X.61.3.192

Molenberghs, P., Trautwein, F.M., Bockler, A., Singer, T. & Kanske, P. (2016) Neural Correlates of Metacognitive Ability and of Feeling Confident: A Large-Scale fMRI Study. *Social, Cognitive and Affective Neuroscience, 11*, 1942-1951. https://doi.org/10.1093/scan/nsw093

Morse, J. M. & Niehaus, L. (2009). *Mixed method design: Principles and procedures.* Left Coast Press.

Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P. & Lee, Y. (2011). Mind your errors. *Psychological Science, 22*(12), 1484-1489. https://doi.org/10.1177/0956797611419520

Moutafi, J., Furnham, A., & Paltiel, L. (2004). Why is conscientiousness negatively correlated with intelligence? *Personality and Individual Differences, 37*(5), 1013–1022. https://doi.org/10.1016/j.paid.2003.11.010

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*, 33-52. http://dx.doi.org/10.1037/0022-3514.75.1.33

Naumann, J. (2019). The skilled, the knowledgeable, and the motivated: Investigating the strategic allocation of time on task in a computer-based assessment. *Frontiers in Psychology, 10,* Article 1429. https://doi.org/10.3389/fpsyg.2019.01429

Neuendorf, K. (2016). *The content analysis guidebook*. SAGE Publications.

Newland, R. (2016). Closing the gap: Understanding two year 10 boys' difficulties with comprehension of Latin stories in a mixed comprehensive school. *Journal of Classics Teaching, 17*(34), 22-30. https://doi.org/10.1017/S2058631016000021O

Nielsen, J. (2000). Designing web usability: The practice of simplicity. New Riders Publishing.

Nielsen, J. & Landauer, T.K., (1993). A mathematical model of the finding of usability problems. *Proceedings of ACM INTERCHI'93 Conference,* 206-213, ACM Press.

Nordman, J. & Adcock, J. (2022). Addressing Low Frustration Tolerance in Students with Learning Disabilities. *Intervention in School and Clinic, 0(0).* https://doi.org/10.1177/10534512221140490

Norman, E. (2020). Why metacognition is not always helpful. *Frontiers in Psychology, 11*, 1537. https://doi.org/10.3389/fpsyg.2020.01537

Nussbaum, D. & Dweck, C. S. (2008). Defensiveness versus remediation: Self theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin, 34*(5), 599-612. https://dx.doi.org/10.1177/0146167207312960

O'Brien, S. (2006). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures, 7*(1), 1–21. https://doi.org/10.1556/Acr.7.2006.1.1

O'Malley, J. & Chamot, A. (1990). *Learning Strategies in Second Language Acquisition* (Cambridge Applied Linguistics). Cambridge University Press.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N. & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and policy in mental health*, *42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Papadopoulos, D. (2020). Effects of a social-emotional learning-based program on self-esteem and self-perception of gifted kindergarten students: A pilot study. *Journal for the Education of Gifted Young Scientists, 8*(3), 1275-1290. https://doi.org/10.17478/jegys.779438

Park, J., Ayduk, Ö. & Kross, E. (2016). Stepping back to move forward: Expressive writing promotes self-distancing. *Emotion, 16*(3), 349–364. https://doi.org/10.1037/emo0000121

Park, B., Korbach, A. & Brünken, R. (2020). Does thinking-aloud affect learning, visual information processing and cognitive load when learning with seductive details as expected from self-regulation perspective? *Computers in Human Behavior, 111,* Article 106411. https://doi.org/10.1016/j.chb.2020.106411

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications.

Patton, M.Q. (2002). *Qualitative research and evaluation methods.* (3rd ed). SAGE Publications.

Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science, 5*(5), 241-248. https://doi.org/10.1111/j.1467-9280.1994.tb00620.x

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18,* 315-341. http://dx.doi.org/10.1007/s10648-006-9029-9

Pekrun, R. & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement*, (259–282). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7_12

Peterson, J. S. (2015). School counselors and gifted kids: Respecting both cognitive and Affective. *Journal of Counselling & Development, 93,* 53-162. https://doi.org/10.1002/j.1556-6676.2015.00191.x

Peterson, J. S. (2018). Counseling gifted children and teens. In S. I. Pfeiffer, E. Shaunessy-Dedrick, & M. Foley-Nicpon (Eds.) *APA handbook of giftedness and talent,* (511-527). American Psychological Association.

Pfeiffer, S. I. & Stocking, V. B. (2000). Vulnerabilities of academically gifted students. *Special Services in the Schools, 16*(1-2), 83-93. https://doi.org/10.1300/J008v16n01_06

Pickard, H. (2003). Emotions and the problem of other minds. In A. Hatimoysis (ed.), *Royal Institute of Philosophy Supplement*, (87-103). Cambridge University Press.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation,* (451–529). Academic Press.

Powers, W. T. (2005). *Behaviour: The control of perception, 2nd edition*. Benchmark Publications.

Pressley, M. & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Erlbaum.

Quihuis, G., Bempechat, J., Jimenez, N. & Boulay, B. (2002). Implicit theories of intelligence across academic domains: A study of meaning making in adolescents of Mexican descent. *New Directions for Child and Adolescent Development, 96,* 87-100. https://doi.org/10.1002/cd.45

Rammstedt, B., Grüning, D. J. & Lechner, C. J. (2022). Measuring Growth Mindset. Validation of a Three-Item and a Single-Item Scale in Adolescents and Adults. *European Journal of Psychological Assessment*. Advance online publication. https://doi.org/10.1027/1015-5759/a000735

Ransdell, S., Levy, C. M. & Kellogg, R. T. (2002). The structure of writing processes as revealed by secondary task demands. *L1 –Educational Studies in Language and Literature, 2*(2), 141-163. https://doi.org/10.1023/A:1020851300668

Reis, S. & Renzulli, J. S. (2004). Current research on the social and emotional development of gifted and talented students: Good news and future possibilities. *Psychology in the Schools, 41*(1), 119-130. https://doi.org/10.1002/pits.10144

Reis, S. M., Sullivan, E. E. & Renzuli, J. S. (2021). Characteristics of gifted learners: Varied, diverse, and complex. In Karnes, F.A., & Bean, S.M. (Eds.). *Methods and Materials for Teaching the Gifted,* (4th ed). Routledge. https://doi.org/10.4324/9781003236603

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*(3), 180-184. https://doi.org/10.1177/003172171109200821

Renzulli, J. S. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness,* 53–92. Cambridge University Press.

Renzulli, J. S. (2016). The three-ring conception of giftedness: A developmental model for promoting creative productivity. In S. M. Reis (Ed.), *Reflections on gifted education: Critical works by Joseph S. Renzulli and colleagues,* (55–90). Prufrock Press Inc..

Renzulli, J. S. & Reis, S. M. (2018). The three-ring conception of giftedness: A developmental approach for promoting creative productivity in young people. In: S. I. Pfeiffer, E. Shaunessy-Dedrick, & M. Foley-Nicpon (Eds.), *APA Handbook of Giftedness and Talent,* (185-199). American Psychological Association. https://doi.org/10.1037/0000038-012

Robertson-Kraft, C. & Duckworth, A. L (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College Record, 116*(3), 1–27. https://doi.org/10.1177/016146811411600306

Rogiers, A., Merchie, E. & Van Keer, H. (2020). What they say is what they do? Comparing task-specific self-reports, think-aloud protocols, and study traces for measuring secondary school students' text-learning strategies. *European Journal of Psychology of Education, 35*, 315–332. https://doi.org/10.1007/s10212-019-00429-5

Romero, C., Master, A., Paunesku, D., Dweck, C. S. & Gross, J. J. (2014). Academic and emotional functioning in middle school: The role of implicit theories. *Emotion*, *14*(2), 227-234. https://doi.org/10.1037/a0035490

Ruiter, S. A. J., Hurks, P. P. M. & Timmerman, M. E. (2017). IQ-score is dringend aan modernisering toe: Naar een nieuwe interpretatie en classificatie van de geschatte intelligentie. [IQ score is in urgent need of modernization: Towards a new interpretation and classification of estimated intelligence.]. *Kind en Adolescent Praktijk*, *16*(1), 16-23. https://doi.org/10.1007/s12454-017-0005-y

Russo, E. J., Johnson, E. J. & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*(6), 759-769. https://doi.org/10.3758/BF03202637

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers*. (3rd ed.) SAGE publications.

Saldana, J. (2021). *The Coding Manual for Qualitative Researchers*. (4th ed.). SAGE Publications.

Sarkissian, J. (2008). Points of emphasis: Some observations based on translation study. In J. Sarkassian (ed.) *AP® Latin: How grammar contributes to literal translation and reading comprehension,* (91-106). The College Board.

Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal, 100*(1), 231-254. https://doi.org/10.1111/modl.12313

Saunders, S. A. (2013). *The impact of a growth mindset intervention on the reading achievement of at-risk adolescent students*. University of Virginia.

Schellings, G. L. M. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning*, *6*(2), 91-109. https://doi.org/10.1007/s11409-011-9069-5

Schellings, G. L. M., Van Hout-Wolters, B. H. A. M., Veenman, M. V. J. & Meijer, J. (2013). Assessing metacognitive activities: The in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education, 28*(3), 963–990. https://doi.org/10.1007/s10212-012-0149-y.

Schilperoord, J. (2001). On the Cognitive Status of Pauses in Discourse Production. In G. Rijlaarsdam (Series Ed.), T. Olive & M. Levy (vol. Eds), *Studies in Writing, vol. 10: Contemporary tools and techniques for studying writing,* (61–87). Kluwer Academic Publishers.

Schnotz, W. & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469–508. https://doi.org/10.1007/s10648-007-9053-4

Schouteten, J., Verwaeren, J., Lagast, S., Gellynk, X. & De Steur, H. (2018). Emoji as a tool for measuring children's emotion when tasting food. *Food quality and preference, 68,* 322-331. https://doi.org/10.1016/j.foodqual.2018.03.005

Schunk, D., Meece, J. & Pintrich, P. (2013). *Motivation in Education: Pearson New International Edition*. Pearson Education Ltd.

Seymour, K. E. & Miller, L. (2017). ADHD and depression: the role of poor frustration tolerance. *Current developmental disorders reports*, *4*, 14-18. https://doi.org/10.1007/s40474-017-0105-2

Silver, C. & Lewins, A. (2014). *Using software in qualitative research*. SAGE Publications. https://doi.org/10.4135/9781473906907

Silverman, L. K. (2002). Asynchronous development. In: M. Neihart, S. Reis, N Robinson & S. Moon, (Eds). *The Social and emotional development of gifted children: What do we know? National Association for Gifted Children*. Prufrock Press, 31-37.

Smith, M. (2015). From adversity to buoyancy: Reconceptualising academic resilience. *The Psychologist, 28,* 690-695.

Sobocinski, M., Järvelä, S., Malmberg, J., Dindar, M., Isosalo, A. & Noponen, K. (2020). How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning? *Metacognition and Learning*, *15*, 99–127. https://doi.org/10.1007/s11409-020-09224-w

Solomon, L. J. & Rothblum, E. D. (1984). Academic procrastination: Frequency and cognitive behavioral correlates. *Journal of Counselling Psychology, 31*(4), 503-509. https://doi.org/10.1037/0022-0167.31.4.503

Sorrenti, L., Spadaro, L., Mafodda, A. V., Scopelliti, G., Orecchio, S. & Filippello, P. (2019). The predicting role of school Learned helplessness in internalizing and externalizing problems. An exploratory study in students with Specific Learning Disorder. *Mediterranean Journal of Clinical Psychology*, *7*(2). https://doi.org/10.6092/2282-1619/2019.7.2035

Spinath, B. & Stiensmeier-Pelster, J. (2001). Implicit theories about the malleability of intelligence and ability. *Psychologische Beiträge, 43*, 53-76. https://doi.org/10.1016/j.cedpsych.2016.01.002

Stoeber, J. & Hotham, S. (2013). Perfectionism and social desirability: Students report increased perfectionism to create a positive impression. *Personality and Individual Differences, 55*(5), 626-629. https://doi.org/10.1016/j.paid.2013.04.023

Straka, O., Portešová, Š., Halámková, D. & Jabůrek, M. (2021). Metacognitive monitoring and metacognitive strategies of gifted and average children on dealing with deductive reasoning task. *Journal of eye movement research*, *14*(4), 1-21. https://doi.org/10.16910/jemr.14.4.1

Subotnik, R. F., Olszewski-Kubilius., P. & Worrell, F, (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest, 12*(1), 3-54. https://doi.org/10.1177/1529100611418056

Sullivan, L. E. (Ed.) (2009). *The SAGE glossary of the social and behavioural sciences, (Vols. 1-3)*. SAGE Publications. https://doi.org/10.4135/9781412972024

Sullivan, P. A., Askew, M., Cheeseman, J., Clarke, D. M., Mornane, A., Roche, A. & Walker, N. (2015). Supporting teachers in structuring mathematics lessons involving challenging tasks. *Journal of Mathematics Teacher Education, 18*(2), 123-140. https://doi.org/10.1007/s10857-014-9279-2

Sullivan, G. M. & Feinn, R. (2012). Using Effect Size-or Why the P Value Is Not Enough. *Journal of graduate medical education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Sun, S. (2011). Think-aloud-based translation process research: Some methodological considerations. *Meta, 56*(4), 928–951. https://doi.org/10.7202/1011261ar

Sun, R. & Mathews, R. (2012). Implicit cognition, emotion, and meta-cognitive control. *Mind & Society, 11,* 107-119. https://doi.org/10.1007/s11299-012-0101-5.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Taber, K. S. (2014). Methodological issues in science education research: A perspective from the philosophy of science. In: M. R. Matthews (Ed.) *International Handbook of Research in History Philosophy and Science Teaching, 3,* (1839-1893). Springer Netherlands.

Taber, K. S. (2018). The use of Cronbach's alpha when developing research instruments in science education. *Research in Science Education, 48*, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches. Applied Social Research Methods Series, Vol. 46*. SAGE Publications.

Ten Have, P. (2007). *Doing conversation analysis: A practical guide*. (2nd ed.). SAGE Publications.

Tilley, S. A. (2003a). "Challenging" Research Practices: Turning a Critical Lens on the Work of Transcription. *Qualitative Inquiry*, *9*(5), 750–773. https://doi.org/10.1177/1077800403255296

Tilley, S. A. (2003b). Transcription work: Learning through co-participation in research. *QSE: International Journal of Qualitative Studies in Education*, *16*(6), 835-851.

https://doi.org/10.1080/09518390310001632171
Tilley S. A. & Powick K. D. (2002). Distanced data: Transcribing other people's research tapes. *Canadian Journal of Education*, *27*(2/3), 291–310. https://doi.org/10.2307/1602225.
Tomlinson, B. (1984). Talking about the Composing Process: The Limitations of Retrospective Accounts. *Written Communication 1*(4), 429-445. https://doi.org/10.1177/0741088384001004003
Tuleja, E. A., Beamer, L., Shum, C. & Chan, E. K. Y. (2011). Designing and developing questionnaires for translation tutorial. *IEEE Transactions on Professional Communication, 54*(4), 392-405. https://doi.org/10.1109/TPC.2011.2172834
Tulis, M. & Fulmer, S. M. (2013). Students' motivational and emotional experiences and their relationship to persistence during academic challenge in mathematics and reading. *Learning and Individual Differences, 27,* 35–46. https://doi.org/10.1016/j.lindif.2013.06.003
Van de Velde, S., Van Keer, H., Schellings, G. & Van Hout-Wolters, B. (2015). Using think-aloud protocol analysis to gain in-depth insights into upper primary school children's self-regulated learning. *Learning and Individual Differences, 43,* 11–30. https://doi.org/10.1016/j.lindif.2015.08.027
Van den Bergh, H. & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process and relationships with text quality. *Educational Psychology, 21*(4), 373–385. https://doi.org/10.1080/01443410120090777
Van der Plaat, A. (2016). *Klassieke talen: Vakspecifieke trendanalyse 2015*. [Classical languages: Subject specific trend analysis 2015]. SLO. Retrieved from: https://www.slo.nl/@4493/klassieke-talen
Van der Stel, M. & Veenman, M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences*, *18*(1)*,* 128-134. https://doi.org/10.1016/j.lindif.2007.08.003
Van Dijk, W. W., Zeelenberg, M. & van der Pligt, J. (1999). Not having what you want versus having what you don't want: The impact of the type of negative outcome on the experience of disappointment and related emotions. *Cognition and Emotion*, *13*, 129-148. https://doi.org/10.1080/026999399379302
Van Gerven, E. (2002). Mag het een niveautje hoger zijn? Schoolbrede organisatie van verrijkingsonderwijs [Can it be a level higher? School-wide organization of enrichment education]. *Talent, 4*(6), 9-11.
Van Houdt, T. (2008). The Strategic Reading of Latin (and Greek) Texts: A Research-Based Approach. In (B. Lister, Ed.) *Meeting the Challenge. International Perspectives on the Teaching of Latin*, 54–70. Cambridge University Press.
Van Krieken, R. (1982). Vertalen en Begrijpen [Translating and understanding]. *Toegepaste Taalwetenschap in Artikelen* , *13*(2), 128–46. https://doi.org/10.1075/ttwia.13.09kri
Van Someren, M., Barnard, Y. F. & Sandberg, J. (1994). *The think-aloud method: a practical approach to modelling cognitive.* AcademicPress.
Veenman M. V. J. (2008). Giftedness: Predicting the speed of expertise acquisition by intellectual ability and metacognitive skillfulness of novices. In M. F. Shaughnessy, M. V. J. Veenman & C. Kleyn-Kennedy (Eds.) *Meta-cognition: A recent review of research, theory, and perspectives*, (207-220). Nova Science Publishers.
Veenman, M. V. J. (2011). Alternative Assessment of Strategy Use with Self-report Instruments: A discussion. *Metacognition and Learning*, *6*, 205-211. https://doi.org/10.1007/s11409-011-9080-x
Veenman, M. V. J. (2013). Training metacognitive skills in students with availability and production deficiencies. In H. Bembenutty, T. Cleary, & A. Kitsantas (Eds.), *Applications of Self-Regulated*

*Learning across diverse disciplines: A tribute to Barry J. Zimmerman,* 299-324. Information Age Publishing.

Veenman, M. V. J. (2015). Het herkennen en instrueren van metacognitieve vaardigheden *[Recognizing and instructing metacognitive skills].* Presentation Symposium Talentontwikkeling in de 21ste eeuw.

Veenman, M. V. J. (2017). Learning to self-monitor and self-regulate. In R. Mayer, & P. Alexander (Eds.), *Handbook of research on learning and instruction, 2nd revised edition,* (233-257). Routledge.

Veenman, M. V. J., Elshout, J. J. & Groen, M. G. M. (1993). Thinking aloud: Does it affect regulatory processes in learning? *Tijdschrift Voor Onderwijsresearch, 18*(6), 322–330.

Veenman, M. V. J., Elshout, J. J. & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction, 7*(2), 187-209. https://doi.org/10.1016/S0959-4752(96)00025-4

Veenman, M. V. J., Kerseboom, L. & Imthorn, C. (2000). Test anxiety and metacognitive skilfulness: Availability versus production deficiencies. *Anxiety, Stress & Coping: An International Journal, 13*(4), 391–412. https://doi.org/10.1080/10615800008248343

Veenman, M. V. J., Kok, R. & Blöte, A. W. (2005). The Relation between Intellectual and Metacognitive Skills in Early Adolescence. *Instructional Science, 33*, 193-211. http://dx.doi.org/10.1007/s11251-004-2274-8

Veenman, M. V. J., Prins, F. J. & Verheij, J. (2003). Learning styles: self-reports versus thinking aloud measures. *British Journal of Educational Psychology, 73*, 357–372. https://doi.org/10.1348/000709903322275885

Veenman, M. V. J. & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15*(2), 159–176. https://doi.org/10.1016/j.lindif.2004.12.001

Veenman, M. V. J. & Van Hout-Wolters, B. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation, 12*(3), 209-237. https://doi.org/10.1080/13803610500479991.

Veenman, M. V. J., Wilhelm, P. & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*(1), 89–109. https://doi.org/10.1016/j.learninstruc.2003.10.004

Veine, S., Kalvig Anderson, M., Haugland Andersen, N., Espenes, T. C., Søyland, T. B., Wallin, P. & Reams, J. (2020). Reflection as a core student learning activity in higher education - Insights from nearly two decades of academic development, *International Journal for Academic Development*, *25*(2), 147–161. https://doi.org/10.1080/1360144X.2019.1659797

Vermunt, J. D. (1989). The interplay between internal and external regulation of learning, and the design of process-oriented instruction. Paper presented at the Third Conference of the European Association for Research on Learning and Instruction, Madrid, Spain.

Vermunt, J. D. (1992). *Leerstijlen en sturen van leerprocessen in het hoger onderwijs - Naar procesgerichte instructie in zelfstandig denken* [Learning styles and guiding learning processes in higher education -Towards process-oriented instruction in independent thinking]. Swets & Zeitlinger.

Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education*, *31*. 25-50. https://doi.org/10.1007/BF00129106.

Vermunt, J. D., Vermetten, Y. J. (2004). Patterns in Student Learning: Relationships Between Learning Strategies, Conceptions of Learning, and Learning Orientations. *Educational Psychology Review, 16*, 359–384. https://doi.org/10.1007/s10648-004-0005-y

Verrier, D., Johnson, S. & Reidy, L. (2018). The teacher academic buoyancy scale: Is it possible to keep tabs on students' academic buoyancy? *International Journal of Assessment Tools in Education, 5*(4), 659-667. https://doi.org/10.21449/ijate.463771

Wade, S. E. (1990). Using think alouds to assess comprehension. *The Reading Teacher*, *43*(7), 442-451.

Wallace, B. (2000). *Teaching the very able child: Developing a policy and adopting strategies for provisions.* David Fulton Publishers.

Walonoski, J. A. & Heffernan, N. T. (2006). Detection and analysis of off-task gaming behaviour in intelligent tutoring systems. In M. Ikeda, K. Ashlay & T.-W. Chan (Eds.), *Intelligent Tutoring Systems: 8th International Conferences, ITS 2006, Jhongli, Taiwan,* (382-391). Springer.

Weiner, B. (2018). The legacy of an attribution approach to motivation and emotion: A no-crisis zone. *Motivation Science, 4*(1), 4–14. https://doi.org/10.1037/mot0000082

Werner, O. & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In: R. Naroll & R. Cohen (Eds.), *A handbook of method in cultural anthropology,* 398-420. American Museum of Natural History.

Wilkinson, C., Carter, B., Satchwell, C. & Bray, L. (2022). Using methods across generations: researcher reflections from a research project involving young people and their parents, *Children's Geographies, 20*(5), 648-660.

Williamson, J. M. (2018). Teaching to individual differences in science and engineering librarianship. Chandos Publishing. https://www.doi.org/10.1016/B978-0-08-101881-1.00009-1

Wilson, T. D. (1994). The Proper Protocol: Validity and Completeness of Verbal Reports. *Psychological Science*, *5*(5), 249-252. https://doi.org/10.1111/j.1467-9280.1994.tb00621.x

Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance*, 36–48. Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-3

Winsor, D. L., & Mueller, C. E. (2020). Depression, suicide, and the gifted student: A primer for the school psychologist. *Psychology in the Schools*, *57*(10), 1627-1639. https://doi.org/10.1002/pits.22416

Wright, K. A., Lam, D. H. & Brown, R. G. (2009). Reduced approach motivation following nonreward: Extension of the BIS/BAS scales. *Personality and Individual Differences, 47*(7), 753-757. https://doi.org/10.1016/j.paid.2009.06.015

Yeager, D. S. & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*, 302-314. https://doi.org/10.1080/00461520.2012.7228-5

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C.P, Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Man Yang, S., Carvalho, C. M., Hahn, P.R., Gopalan, M., Mhatre, P., Ferguson, R., Duckworth, A. L. & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364-369. https://doi.org/10.1038/s41586-019-1466-y

Young, K. (2005). Direct from the source: the value of 'think-aloud' data in understanding learning. *Journal for Educational Enquiry, 6*(1), 19-33.

Young, A. E. & Worrell, F. C. (2018). Comparing meta- cognition assessments of mathematics in academically talented students. *Gifted Child Quarterly*, *62*(3), 259-275. https://doi.org/10.1177/0016986218755915

Zhang, L. J. & Zhang, D. (2019). Think-aloud protocols. In J. McKinley & H. Rose (Eds.) *The Routledge handbook of research methods in applied linguistics,* (302-331). Routledge.

Zhang, J. & Zhang, L. J. (2022). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning.*

Advance online publication. https://doi.org/10.1080/09588221.2022.2069822

Ziegler, A. & Heller, K. A. (2000). Conceptions of giftedness from a meta-theoretical perspective. In K. A. Heller, F. J. Mönks, R. J. Sternberg and R.F. Subotnik (Eds.), *International handbook of giftedness and talent*, (3-21). Elsevier Science Ltd.

Ziegler, A. & Phillipson, S. N. (2012). Towards systemic theory of gifted education. *High Ability Studies, 23*(1), 3-31. https://doi.org/10.1080/13598139.2012.679085

Ziegler, A. & Stoeger, H. (2010). Research on a modified framework of implicit personality theories. *Learning and Individual Differences, 20*(4), 318-326. https://doi.org/10.1016/j.lindif.2010.01.007

Zimmerman, B. J. (2000). Attainment of self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation, (13-39). Academic Press.

Zimmerman, B. J. & Schunk, D. H. (Eds.). (2011). *Handbook of self-regulation of learning and performance.* Routledge/Taylor & Francis Group.

Zonnenberg, L. (2022). *Situatieanalyse voltijd primair onderwijs (PO) voor hoogbegaafden (HB) in Nederland* [Situation analysis of full-time gifted education in the Netherlands]. https://hbscholen.nl/rapport/

Zsu, M. & Urhahne, D. (2014). Assessing teachers' judgements of students' academic motivation and emotions across two rating methods. *Educational Research and Evaluation: An International Journal on Theory and Practice, 20*(5), 411-427. https://doi.org/10.1080/13803611.2014.964261

AUTHOR INDEX

## SUMMARY

## DEALING WITH SETBACKS

Worldwide, gifted adolescents are at significant risk of underachieving or even dropping-out of school. The Netherlands is no exception, with many of its gifted students leaving school prematurely. However, at the same time there are also gifted students who are academically thriving. A psychological phenomenon that might go some way in explaining this difference is academic buoyancy. Academic buoyancy refers to the ability to deal with daily setbacks that are inherent to school. When a student is less able to deal with these setbacks such as being asked to perform a too difficult task, this can cause academic disengagement.

Existing studies relied on self-report to determine to what extent participants were academically buoyant. We, however, set out to create a task-based measure of academic buoyancy. To do so, we focused on the Latin translation domain. This was a suitable domain to investigate academic buoyancy, as Latin texts are of a heterogeneous in their complexity. The fact that a particularly difficult sentence can be followed by a much easier one, provided us with the opportunity to create a task with differing complexity which seemed 'normal' to the participants, yet enabled us to study academic buoyancy by triggering a setback within the task. Hence, we created the Latin Buoyancy Task.

The design of the Latin Buoyancy Task was the central topic of the first part of Chapter 1. We presented the underlying design principles and the subsequent design cycles of the Latin Buoyancy Task which led to the instrument itself. The instrument existed of three sequential parts: two relatively easy translation tasks which functioned as the pre- and post-measurements. A visually similar task, yet in reality a task which was untranslatable, served as the setback and the intervention. This instrument enabled us to compare the participants' translation accuracy and processes before and after the setback, thus operationalising academic buoyancy.

In the remaining three parts of Chapter 1, we introduced the other four instruments which we created for our study. The first of these was a task that we used to purposefully select participants for the study. This instrument was also created in the form of a design study. This design study was the central topic of the second part of the first chapter. In Chapter 1 we also introduced our Dutch versions of questionnaires to measure mindset preference and frustration tolerance. Previously, no Dutch version of these questionnaires had been statistically validated. Finally, we presented the Emoji Chart, which aimed at gathering data regarding the affective state of the participants at set intervals during the thinking aloud.

In Chapter 2 we centred our attention on three methodological aspects of the study. First, we addressed how we collected our data. We particularly focussed on why we made use of thinking aloud to collect the data at the heart of our study. We also dwelt upon the advantages of combining concurrent and retrospective thinking aloud. Then we shared our considerations regarding the operationalisation of giftedness. By focussing on Latin as the domain and schools with the *Begaafdheidsprofielschool* label, we operationalised giftedness, without being reliant on at the time, out-dated Dutch IQ-tests. By following this method, we hoped to construe understanding of academic buoyancy in a specific yet diverse group of gifted students. Finally, in Chapter 2 we explained why extreme sampling fitted our research purposes and which procedures we followed to select our participants. This led to the introduction of 16 selected participants.

In Chapter 3 we set the development and implementation of the 'main coding scheme' and its derived 'contextualizing coding scheme' forth. Each of these schemes was developed in multiple stages, which in themselves could be categorized as literature-driven and data-driven. The main coding scheme consisted of 32 codes and aimed at thoroughly coding the participants' translation process. This coding scheme was particularly suited for quantitative analysis of the concurrent thinking aloud. Due to the many codes and its not including the data from the retrospective thinking aloud, it was less suitable for our qualitative analysis, which did include the retrospective thinking aloud. Therefore, we also created the contextualizing coding scheme, which was derived from the main coding scheme. This coding scheme was shorter and allowed for more room in the coding for affection, based on context and the data gathered via the retrospective thinking aloud. By making use of both the main coding scheme and the contextualized coding scheme we were thus both able to analyse the data qualitatively and quantitatively.

The end of Chapter 3 formed a breakpoint in the dissertation. Where the previous chapters concentrated on the design and research methodology, the following chapters addressed the data analysis itself and the accompanying research questions.

Chapter 4 aimed at answering two research questions via a quantitative analysis: [1] 'to what extent are the participants translation quality and processes affected by the setback?' and [2] 'to what extent is that the effect of the setback on the students' translation quality and processes moderated by their translation proficiency, mindset preference and frustration tolerance?'. To answer these questions regarding translation quality, we created nested models. Using these models, we first compared and contrasted the accuracy scores of the pre-setback translations to those of the post-setback. We found that in general, there was a significant decline in accuracy after the setback task, despite some participants translation quality actually improving. When including the learner variables (translation proficiency, mindset preference and frustration tolerance), we found that when the translation proficiency was higher, the decline in accuracy was larger. This was also the case when the participants' mindset preference leant more towards a fixed mindset. For frustration tolerance, we found no significant interaction effect.

When focussing on the participants' translation processes, we found that after the setback task, they persevered less and displayed less implicit metacognitive behaviour. Again, the extent of the effect of the setback task on their translation process was moderated by the extent of the participants' proficiency and their mindset preference: when the participants were more proficient or more inclined towards a fixed mindset the effect of the setback task was larger. Having found effects of the setback task on both the participants' quality of translation and their translation processes, we wanted to deepen our understanding of these effects. Moreover, we concluded that the data from the setback task itself might provide further insights into how the participants were affected by the setback task.

Hence Chapter 5 concentrated on the setback task itself and included a qualitative analysis. Our objective was to describe to what extent translation proficiency and mindset preference accounted for differences between the participants' display of metacognitive and affective strategies throughout the setback task. To do so, we created visualisations of the participants' translation processes using the contextualised coding scheme. We then made two comparisons.

First, we compared and contrasted the translation processes of the participants with the highest translation proficiency to those of the participants with the lowest proficiency. We observed that negativity occurred sooner in high proficient participants, alongside less indications of both persevering and grit compared to the low proficient participants. Secondly, we contrasted the translation processes of the participants with the most outspoken fixed mindset preference to those of the participants with the most outspoken growth preference. In that case we found that avoidance strategies and negativity only occurred in the participants with a fixed mindset and that motivation loss occurred sooner than in the growth mindset participants. Further analysis of particularly the differences between high and low proficient participants and the inclusion of academic buoyancy was recommended as a next step.

In Chapter 6 we, therefore, created four strategy profiles to ascertain how the learning strategies and patterns of academic buoyant participants compared to those of non-academic buoyant participants when they were faced with the Latin Buoyancy Task (i.e. a task of heterogenous complexity). To create the profiles, we selected four participants each representing one quadrant:

1. High academic buoyant and high proficient at translating Latin;
2. Low academic buoyant and high proficient at translating Latin;
3. High academic buoyant and low proficient at translating Latin;
4. Low academic buoyant and low proficient at translating Latin.

We followed these participants throughout the Latin Buoyancy Task and including their retrospective thinking and the data from the Emoji Chart in our qualitative analysis. We found that the quality of the between the metacognitive and cognitive activity was highest in the high proficient participants and that they relied on bottom-up strategies. For academic buoyancy, we concluded that high buoyant participants were more flexible in their displayed strategies and that that they demonstrated more negativity and avoidance strategies compared to the low academic buoyant participants.

Despite growing educational and political awareness of the problems Dutch gifted students face, research remains scarce and the effect of academic buoyancy on students' task attainment was yet to be investigated in the Netherlands. With this dissertation we filled in part of this void. Our main objective was to establish how a lack of academic buoyancy in gifted students was reflected in their task processes. We also aimed to determine to what extent the effect of a too difficult task on their accuracy and translation processes was moderated by the learner variables translation proficiency, mindset preference and frustration tolerance. On the basis of the studies included in this dissertation, we drew five main conclusions regarding academic buoyancy and Latin translation processes of our gifted participants:

1. Academic buoyancy can be studied via a task-based measure.
2. The heterogeneity of Latin texts necessitates academic buoyancy.
3. There was not only an effect of the increased difficulty on the participants' translation accuracy and processes during the setback task, but for some participants this effect was prolonged.
4. When the participants were either high proficient or more inclined towards a fixed mindset preference, the effect of the setback on the participants' translation accuracy and processes was largest.
5. Throughout the task, the participants' affective strategies underlied the cognitive and metacognitive strategies they used and to what extent they bounced back from the setback task.

With the completion of this dissertation we hope to contribute to our knowledge on gifted students in the Netherlands. We conclude that if fewer gifted Latin students are to underachieve, Latin teachers must pay more attention to the translation process as a whole, notwithstanding the emotions that are paired with learning.

# SAMENVATTING

## OMGAAN MET TEGENSLAGEN

Hoogbegaafde jongeren lopen wereldwijd risico op onderpresteren. Dit kan leiden tot tijdelijke of permanente schooluitval. Ook Nederlandse hoogbegaafde jongeren hebben met dit probleem te maken. Er zijn echter ook hoogbegaafde jongeren die op school wel succesvol zijn. Het verschil tussen deze twee groepen kan mogelijk deels verklaard worden door de mate van academische veerkracht, wat in de literatuur bekend staat als *academic buoyancy. Academic buoyancy* verwijst naar het vermogen van een leerling om kortdurende schoolse tegenslagen, zoals lastige opdrachten of tegenvallende cijfers, het hoofd te bieden. Wanneer een leerling moeite heeft om met zulke tegenslagen om te gaan, kan dit leiden tot verlies van motivatie. Het effect van *academic buoyancy* op het taakgedrag van hoogbegaafde jongeren is echter nog niet onderzocht binnen de Nederlandse onderwijscontext.

De onderzoeken die er wel zijn maken voornamelijk gebruik van vragenlijsten om te bepalen in hoeverre hun deelnemers *academically buoyant* waren. Wij hebben er daarentegen voor gekozen om academic buoyancy te onderzoeken aan de hand van een schoolse taak om sociaal wenselijke antwoorden te vermijden. Voor een dergelijke taak viel de keuze op het het vertalen van Latijn, vanwege de heterogeniteit van Latijnse teksten. In deze teksten komt het namelijk regelmatig voor dat een complexe zin opgevolgd wordt door een eenvoudigere zin. De variatie binnen teksten die leerlingen gewend zijn om te vertalen bij Latijn, bood ons de mogelijkheid om een taak te ontwerpen, die verschillende niveaus van complexiteit bevatte. Hierdoor konden we een tegenslag binnen de taak isoleren. Dit leidde tot de *Latin Buoyancy Task*.

Het eerste deel van Hoofdstuk 1 staat in het teken van het ontwerproces van de *Latin Buoyancy Task*. In dit hoofdstuk presenteren we de opeenvolgende ontwerpcycli, om tot deze taak te komen. Het instrument bestond uit drie delen: twee relatief eenvoudige taken die functioneerden als voor- en nameting. De middelste taak functioneerde als interventie. Op het zicht leek deze taak op de andere twee taken. In werkelijkheid was de taak onvertaalbaar. Via de onoplosbaarheid van deze taak bootsten we een tegenslag na. Met behulp van de Latin Buoyancy Task konden we zowel de correctheid van de vertaling als het vertaalproces van de deelnemers na de tegenslag vergelijken met dat van ervoor. Op deze manier konden we *academic buoyancy* operationaliseren aan de hand van een schoolse taak.

In de overige delen van Hoofdstuk 1 introduceren we de overige vier instrumenten die we hebben ontworpen voor dit onderzoek. Het ontwerp van het selectieinstrument stond centraal in het tweee deel van het eerste hoofdstuk. We presenteren vervolgens onze Nederlandstalige versies van de vragenlijsten, waarmee we *mindset* voorkeur en frustratietolerantie vaststelden. Aangezien er van deze vragenlijsten nog geen statistisch gevalideerde Nederlandstalige versies waren, hebben we de vragenlijsten zelf vertaald en gevalideerd. Tot slot bespreken we de Emoji Chart. Het doel van dit instrument was om gegevens te verzamelen over de affectieve toestand van de deelnemers op gezette momenten tijdens het uitvoeren van de Latin Buoyancy Task.

In Hoofdstuk 2 zetten we drie methodologische aspecten van het onderzoek uiteen. Eerst leggen we uit op welke manier we onze gegevens hebben verzameld. Hierbij ligt de nadruk op waarom hardop denken een geschikte methode was voor onze dataverzameling. Daarnaast bespreken we de voordelen van het combineren van gelijktijdig retrospectief hardop denken. Vervolgens wordt het ingewikkelde begrip 'hoogbegaafdheid' onder de loep genomen. Voor ons onderzoek operationaliseren we hoogbegaafdheid via het vak Latijn en de Begaafdheidsprofielscholen om niet afhankelijk te zijn van de destijds verouderde Nederlandse IQ-testen. Tot slot leggen we in Hoofdstuk 2 uit waarom het selecteren van onze deelnemers via *extreme sampling* paste bij onze onderzoeksdoelen en introduceren we onze 16 deelnemers.

In Hoofdstuk 3 presenteren we de ontwikkeling en implementatie van de twee codeerschema's waarmee we de hardop denk data hebben geanalyseerd. Deze schema's zijn gedurende meerdere fases ontwikkeld, waarvan sommige gebaseerd waren op literatuur en andere op de verzamelde data. Het hoofdschema bestond uit 32 codes en was gericht op het exhaustief coderen van het vertaalproces van de deelnemers. Dit codeerschema was voornamelijk geschikt voor kwantitatieve analyse. Vanwege de vele codes en omdat het geen gebruik maakte van de data die we via het retrospectief hardop denken hebben opgehaald, was het minder geschikt voor onze kwalitatieve analyse. Daarom hebben we ook het contextualiserende codeerschema samengesteld. Dit schema was een kortere afgeleide van het hoofdschema en bood bovendien meer ruimte om ook affectieve uitspraken en strategieën te coderen. Door gebruik te maken van beide codeerschema's konden we zowel kwantitatieve als kwalitatieve analyses uitvoeren.

Hoofdstuk 3 vormt het sluitstuk van het methodologische deel van dit proefschrift. De data-analyse en bijbehorende onderzoeksvragen staan centraal in Deel II.

Het doel van Hoofdstuk 4 is om via een kwantitatieve analyse twee onderzoeksvragen te beantwoorden: [1] 'in welke mate worden de vertaling en vertaalprocessen van de deelnemers beïnvloed door de tegenslag?' en [2] 'in welke mate wordt het effect van de tegenslag op de vertaling en vertaalprocessen gemodereerd door hun vertaalvaardigheid, mindset voorkeur en frustratietolerantie?'. Om deze vragen te beantwoorden maken we gebruik van geneste modellen. Met behulp van deze modellen vergelijken we de kwaliteit van de vertalingen van vóór de tegenslag vertalingen met die van erna. We stellen vast dat er over het algemeen een aanzienlijke achteruitgang in kwaliteit was na de tegenslagtaak, hoewel de vertaalkwaliteit van enkele deelnemers verbetert. Wanneer we rekening houden met de leerlingkenmerken (vertaalvaardigheid, mindset voorkeur en frustratietolerantie), ontdekken we dat de afname in vertaalkwaliteit groter is, als de vertaalvaardigheid hoger is. Dit geldt ook als de *mindset*-voorkeur van de deelnemers meer neigt naar een *fixed mindset*.

Om die reden concentreren we ons in Hoofdstuk 5 op de tegenslagtaak zelf en voeren we een kwalitatieve analyse uit. Ons doel was om te beschrijven in welke mate vertaalvaardigheid en mindset voorkeur verantwoordelijk waren voor verschillen tussen de deelnemers met betrekking tot het inzetten van metacognitieve en affectieve strategieën tijdens de tegenslagtaak. We creëeren visualisaties van de vertaalprocessen van de deelnemers met behulp van het *contextualising* codeerschema. Vervolgens vergelijken we de visualisaties van de verschillende type deelnemers met elkaar.

Eerst vergelijken we de vertaalprocessen van de meest vertaalvaardige deelnemers met die van de deelnemers met de minste vaardigheid. We zien dat negativiteit eerder en vaker voorkomt bij de meest vaardige deelnemers. Daarnaast zagen we minder aanwijzingen van doorzettingsvermogen en *grit* in vergelijking met de minst vaardige deelnemers. Ten tweede vergelijken we de vertaalprocessen van de deelnemers met de meest duidelijke voorkeur voor een *fixed mindset* met die van de deelnemers met de meest duidelijke voorkeur voor een *growth mindset*. We concludeerden dat vermijdingsstrategieën en negativiteit alleen voorkomen bij de deelnemers met een *fixed mindset* en dat zij eerder hun motivatie verliezen dan deelnemers met een *growth mindset*. Verdere analyse van de verschillen tussen de vaardige en niet-vaardige deelnemers en het opnemen van *academic buoyancy* in de analyse werd aanbevolen als volgende stap.

In Hoofdstuk 6 maken we daarom vier strategieprofielen om vast te stellen hoe de leerstrategieën van *academic buoyant* deelnemers zich verhouden tot die van niet-*academic buoyant* deelnemers wanneer ze worden geconfronteerd met de Latin Buoyancy Taak. Om de profielen te maken, selecteren we vier deelnemers die elk één kwadrant vertegenwoordigden:

1. Veel *academic buoyancy* en hoge vaardigheid in het vertalen;
2. Weinig *academic buoyancy* en hoge vaardigheid in het vertalen;
3. Veel *academic buoyancy* en lage vaardigheid in het vertalen;
4. Weinig *academic buoyancy* en lage vaardigheid in het vertalen;

We volgen deze deelnemers tijdens de Latin Buoyancy Task en nemen hun retrospectief hardop denken en de gegevens van de Emoji-Chart op in onze verdiepende kwalitatieve analyse. We ontdekken dat de kwaliteit van de wisselwerking tussen metacognitieve en cognitieve activiteit het hoogst is bij de meest vaardige deelnemers. Deze deelnemers vertrouwden op bottom-up strategieën. Voor *academic buoyancy* concluderen we dat deelnemers met veel *buoyancy* flexibeler waren in hun getoonde strategieën en dat ze meer negativiteit en vermijdingsstrategieën vertoonden in vergelijking met deelnemers met weinig veerkracht.

Ons voornaamste doel was om te bepalen op welke manieren een gebrek aan *academic buoyancy* zichtbaar wordt in het taakgedrag van hoogbegaafde leerlingen. Daarnaast streefden we ernaar om te achterhalen in hoeverre het effect van een te lastige taak wordt beïnvloed door de mate van vertaalvaardigheid, *mindset* voorkeur en frustratietolerantie. Op basis van de studies die onderdeel uitmaakten van dit proefschrift, trekken we over *academic buoyancy* en de vertaalprocessen van onze hoogbegaafde deelnemers vijf conclusies:

1. *Academic buoyancy* kan via een taak onderzocht worden.
2. De heterogeniteit van Latijnse teksten vereist een zekere mate van *academic buoyancy*.
3. Er is niet alleen een effect van de verhoogde moeilijkheid op de vertaling en de vertaalprocessen van de deelnemers tijdens de tegenslagtaak, maar voor sommige deelnemers is dit effect langdurig.
4. Als de deelnemers of goede vertalers zijn of een sterke voorkeur hebben voor een *fixed mindset*, dan is het effect van de tegenslag op de vertaling en de vertaalprocessen het grootst.
5. Gedurende de hele taak waren de affectieve strategieën van de deelnemers de basis voor de cognitive en metacognitieve strategieën die ze gebruikten en in hoeverre ze herstelden van de tegenslag.

Hoewel er vanuit onderwijsvoorzieningen en de politiek steeds meer aandacht is voor de problemen waarmee Nederlandse hoogbegaafde leerlingen te maken hebben, blijft relevant Nederlands onderzoek naar hoogbegaafde (voortgezet onderwijs) leerlingen nog veelal achterwege. Met het voltooien van dit proefschrift hopen we een bijdrage te leveren aan deze problematiek.

# APPENDICES

APPENDICES: TABLE OF CONTENTS

APPENDIX 1.1

LATIN BUOYANCY TASK & WORDLIST

PRE-SETBACK TASK

*Equus pilo traiectus est.* *Liv. 27.32.5*

.......................................................................................................................................................

*Caesar suos a proelio continebat.* *Caes. DBG. 15.2*

.......................................................................................................................................................

*Veniam ille amori forsitan nostro dat*. *Sen. Ph. 225*

.......................................................................................................................................................

Wordlist

| | |
|---|---|
| a *+abl.* | van |
| amor, amoris | liefde |
| Caesar | Caesar *bekende Romeinse generaal* |
| contineo,continere | weghouden |
| do, dare | geven |
| equus | paard |
| fortisan | misschien |
| ille | hij |
| noster | ons, onze |
| pilus | speer |
| proelium | strijd |
| sum, fui, esse | zijn |
| suus, a, um | *hier:* zijn man |
| traiectus, a, um | doorboren |
| venia | gif, vergiffenis |

## SETBACK TASK

*Hic sine filium longe gravissimum cano.*

..................................................................................................................................................

*In mare habes Delphinus et canorum promittere.*

..................................................................................................................................................

*Cur esse ea bona periit similis copia*?

..................................................................................................................................................

Wordlist

| | |
|---|---|
| bonus, a um | goed |
| cano, canere | zingen |
| canus | hond |
| copia | hulptroep, voorraad |
| cur | waarom |
| delphinus | dolfijn |
| et | en |
| filius | zoon |
| gravis, -is | zwaar |
| habeo. habere | hebben |
| hic | hier |
| in *+abl.* | in |
| is, ea, id | deze |
| longe *adv.* | lang |
| mare, maris | zee |
| pereo, perire | sterven |
| promitto, promittere | beloven |
| similis *+gen*. | gelijkend op |
| sine *+abl.* | zonder |
| sum, fui, esse | zijn |

POST SETBACK TASK

*Vitam iucundissimam vivo.*

....................................................................................................................................................

*Neptunus tridente suo terram percussit.*

....................................................................................................................................................

*Octavianus armis expulit ex urbe collegam.*

....................................................................................................................................................

Wordlist

| | |
|---|---|
| arma | wapen(s) |
| collega | collega |
| ex *+abl.* | uit |
| expello, expuli, expellere | verdrijven |
| iucundus, a um | prettig |
| Neptunus | Neptunus, *de god van de zee* |
| Octavianus | Octavianus, *de eerste keizer van Rome* |
| percutio, percussi, percutere | doen trillen |
| suus, a um | zijn, haar |
| terra | aarde, land |
| tridens, tridentis | drietand |
| urbs, urbis | stad |
| vita | (het) leven |
| vivo, vivere | leven |

## APPENDIX 1.2

## ANSWER KEY TO THE ANAGRAM TASK

| | |
|---|---|
| AKERT: | kreta, raket, kater |
| AETLR: | ratel, alert, alter, later |
| EKLON: | kolen, loken, onkel |
| EKOPR: | kroep, poker, oprek, koper |
| EMNOR: | moren, enorm, romen |
| EOPRT: | poert, roept, troep |
| EOPST: | pesto, sopte, poste, stoep, poets |
| AGKLU: | - |
| BEFOR: | - |
| BEGLO: | globe |
| EIKLP: | - |
| EKLNU: | kleun |
| ENNOX: | xenon |
| HIPTU | - |
| AEGLN: | lange, lagen, angel, algen, nagel |
| AEGRV: | grave, vager, gaver, verga |
| AEMNR: | armen, ramen, maren |
| EEGIN: | genie, eigen, enige |
| EKORT: | koert, korter, koter |
| EKOST: | koest, koets, kotse, koste |
| ENORT: | toren, tenor, toner |

## APPENDIX 1.3

## DUTCH MINDSET AND FRUSTRATION TOLERANCE QUESTIONNAIRE

Lees de onderstaande stellingen. Omcirkel per stelling het nummer dat het best aangeeft in hoeverre jij het met de stelling eens bent. Er wordt hier gevraagd om een mening; er zijn geen goede of foute antwoorden. De antwoorden worden volledig anoniem verwerkt.
Example of the format

1. Ik ben geboren met een bepaalde basisintelligentie en ik kan niet echt iets doen om dit te veranderen.

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| Helemaal mee oneens | Mee oneens | Een beetje mee oneens | Een beetje mee eens | Mee eens | Helemaal mee eens |

All items

1. Ik ben geboren met een bepaalde basisintelligentie en ik kan niet echt iets doen om dit te veranderen.
2. Als ik door omstandigheden een belangrijk doel niet kan halen, vind ik het moeilijk om het toch te blijven proberen.
3. Als ik een uitleg niet snap, haak ik af en let ik niet meer op.
4. Intelligentie is een basisonderdeel van mezelf waar ik weinig aan kan veranderen.
5. Als iets waar ik naar uit kijk niet doorgaat, dan heb ik geen zin of energie om een alternatief te regelen.
6. Als ik een toets heel moeilijk vind, beantwoord ik de vragen maar gewoon niet meer.
7. Als het niet gaat zoals ik wil, kunnen mijn dagelijkse bezigheden me minder schelen.
8. Ik kan nieuwe dingen leren, maar mijn basisintelligentie kan ik niet veranderen.
9. Als ik ergens hard aan heb gewerkt, maar niet de waardering krijg die ik zou verdienen, dan ben ik minder gemotiveerd.
10. Als iets leuks onverwacht toch niet doorgaat, ben ik een tijd lang minder opgewekt.
11. Als ik merk dat mijn huiswerk moeilijker is dan ik verwachtte, word ik chagrijnig.

APPENDIX 2.1

THE THINK ALOUD WARM-UP TASK

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4 | 8 | | 2 | | 9 |
| | | | | 7 | | | 5 | 1 |
| | 8 | 3 | | 2 | | | | |
| | | 4 | | | | | | |
| 7 | 6 | | | | | | | 2 |
| | 5 | | 7 | | 9 | | | |
| | | 7 | | | 5 | 9 | | 4 |
| | | | | | | 5 | | |
| 4 | | | 8 | | | 6 | 7 | |

Source: http://allfreeprintable.com/cont/sdku/pdf/sdku-m-1.pdf

## APPENDIX 4.1

## ESTIMATES OF EFFECTS: FACTORS IN NESTED MODELS

Effect on Accuracy: Learner variable (LV) proficiency. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 1.486 | 0.1 | 18.154 | 14.829 | <.001 |
| Pre-test | 2.274 | 0.039 | 1110.993 | 58.624 | <.001 |
| LV-centred | 0.082 | 0.08 | 18.947 | 1.023 | 0.319 |
| LV-centred in Pre-test | 0.968 | 0.032 | 1109.886 | 30.192 | <.001 |

Effect on Accuracy: Learner variable (LV) Mindset. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 1.535 | 0.266 | 16.528 | 5.76 | <.001 |
| Pre-test | 2.319 | 0.052 | 1108.684 | 44.381 | <.001 |
| LV-centred | 0.026 | 0.226 | 16.577 | 0.116 | 0.909 |
| LV-centred in Pre-test | -0.194 | 0.045 | 1108.561 | -4.315 | <.001 |

Effect on Accuracy: Learner variable (LV) Frustration Tolerance. Model 1

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 1.542 | 0.264 | 16.531 | 5.835 | <.001 |
| Pre-test | 2.295 | 0.052 | 1108.763 | 43.806 | <.001 |

Effect on Task Time: Learner variable (LV) proficiency. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 372,517 | 38,492 | 16,307 | 9,678 | <,001 |
| Pre-test | 226,742 | 5,857 | 1108,458 | 38,712 | <,001 |
| LV-centered | -30,727 | 30,318 | 16,416 | -1,013 | 0,326 |
| LV-centered in Pretest | 14,797 | 4,841 | 1108,284 | 3,057 | 0,002 |

Effect on Task Time: Learner variable (LV)Mindset. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 363.209 | 39.468 | 16.268 | 9.203 | <.001 |
| Pre-test | 234.282 | 5.548 | 1108.351 | 42.227 | <.001 |
| LV-centred | 41.138 | 33.425 | 16.293 | 1.231 | 0.236 |
| LV-centred in Pre-test | -58.524 | 4.766 | 1108.287 | -12.279 | <.001 |

Effect on Task Time: Learner variable (LV) Frustration Tolerance. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 375.669 | 37.334 | 16.327 | 10.062 | <.001 |
| Pre-test | 224.711 | 5.847 | 1108.484 | 38.43 | <.001 |
| LV-centred | 33.324 | 46.723 | 16.42 | 0.713 | 0.486 |
| LV-centred in Pre-test | 35.537 | 7.667 | 1108.661 | 4.635 | <.001 |

Effect on Duration of Activity: Learner variable (LV) proficiency. Model 1

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 18.002 | 1.082 | 16.36 | 16.636 | <.001 |
| Pre-test | -1.394 | 0.177 | 1108.521 | -7.858 | <.001 |

Effect on Duration of Activity: Learner variable (LV) Mindset. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 17.873 | 0.998 | 16.427 | 17.917 | <.001 |
| Pre-test | -1.489 | 0.176 | 1108.555 | -8.446 | <.001 |
| LV-centred | 0.94 | 0.845 | 16.466 | 1.113 | 0.282 |
| LV-centred in Pre-test | 0.783 | 0.151 | 1108.455 | 5.168 | <.001 |

Effect on Duration of Activity: Learner variable (LV) Frustration Tolerance. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 18.123 | 1.076 | 16.34 | 16.839 | <.001 |
| Pre-test | -1.532 | 0.172 | 1108.503 | -8.908 | <.001 |
| LV-centred | -1.373 | 1.347 | 16.437 | -1.019 | 0.323 |
| LV-centred in Pre-test | 2.034 | 0.225 | 1108.687 | 9.023 | <.001 |

Effect on Number of Switches: Learner variable (LV) proficiency. Model 3

| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 22.328 | 2.516 | 16.349 | 8.874 | <.001 |
| Pre-test | 15.867 | 0.408 | 1108.519 | 38.927 | <.001 |
| LV-centred | -2.17 | 1.982 | 16.473 | -1.095 | 0.289 |
| LV-centred in Pre-test | -1.125 | 0.337 | 1108.322 | -3.338 | <.001 |

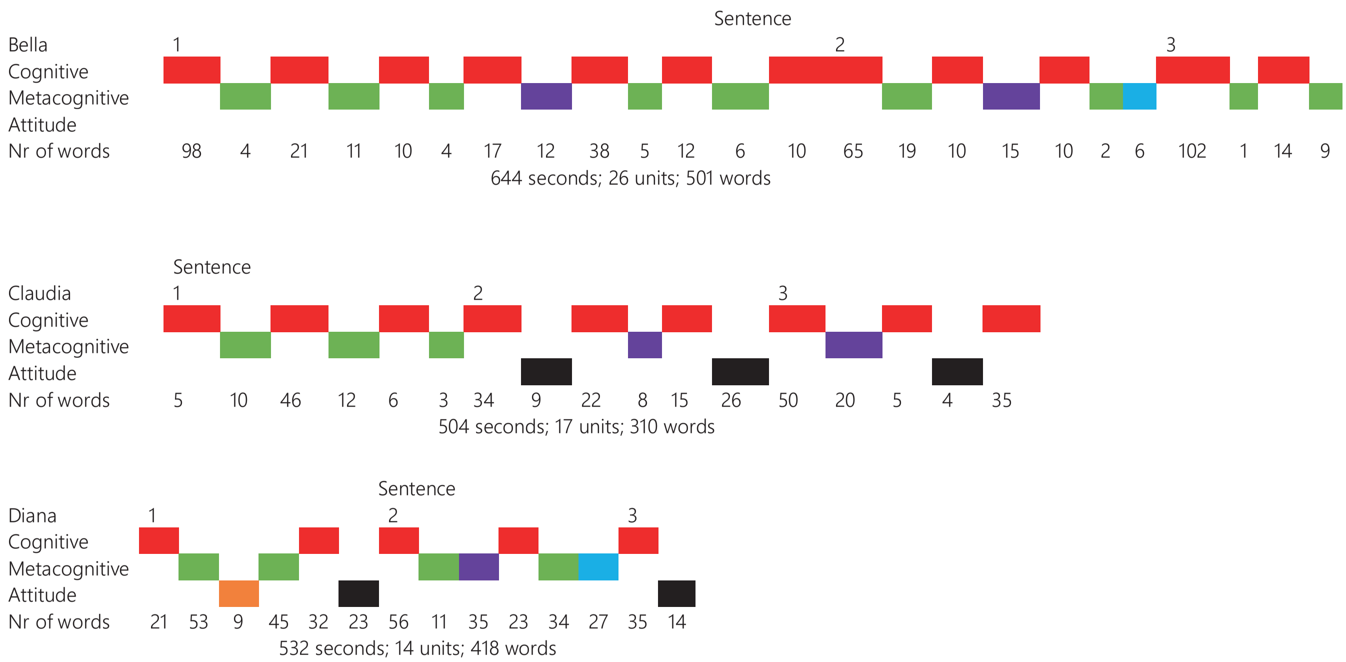
Effect on Number of Switches: Learner variable (LV) Mindset. Model 3

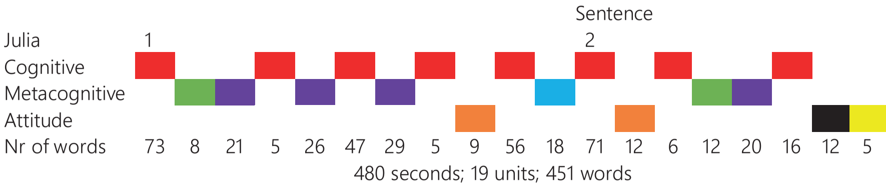
| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 22.039 | 2.487 | 16.31 | 8.863 | <.001 |
| Pre-test | 16.433 | 0.375 | 1108.404 | 43.793 | <.001 |
| LV-centred | -0.314 | 2.106 | 16.338 | -0.149 | 0.883 |
| LV-centred in Pre-test | -4.855 | 0.322 | 1108.331 | -15.062 | <.001 |

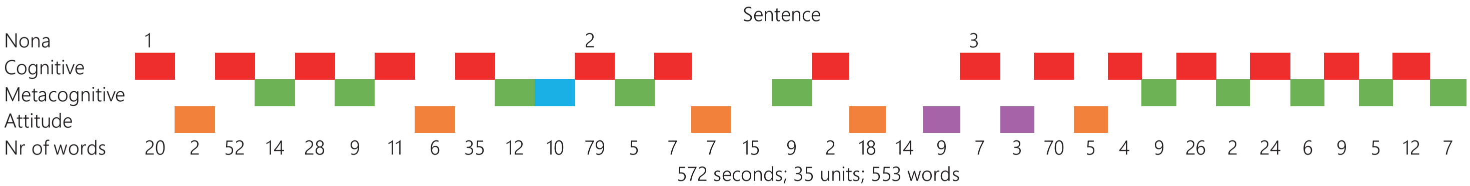
Effect on Number of Switches: Learner variable (LV) Frustration Tolerance. Model 3

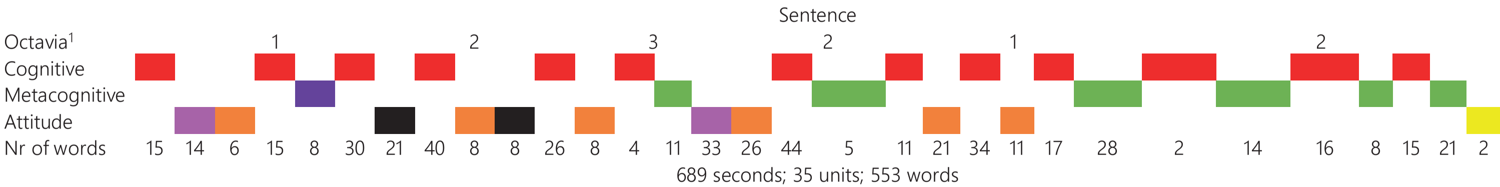
| Parameter | Estimate | *se* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | 22.182 | 2.616 | 16.328 | 8.48 | <.001 |
| Pre-test | 15.914 | 0.41 | 1108.485 | 38.772 | <.001 |
| LV-centred | 3.883 | 3.273 | 16.422 | 1.186 | 0.252 |
| LV-centred in Pre-test | -1.079 | 0.538 | 1108.663 | -2.005 | 0.045 |

# APPENDIX 5.1. PROCESS BARS OF THE REMAINING PARTICIPANTS






[1] *Octavia switches after sentence 3 tot sentence 2, then 1, then 2 again*

## APPENDIX 6. 1

## NARRATIVE OF ANTHONY'S TRANSLATION PROCESS

Anthony was a 15-year-old whose score decreased by five on the Anagram Task after the difficult anagrams. His self-reported mindset was entity (*M* = 2.67), indicating that he would likely be affected negatively by being unable to solve the second task, as he indeed was.

### TRANSLATION PROCESS DURING THE PRE- AND POST-SETBACK TASK

Before starting the think-aloud task Anthony felt happy when asked to fill in the emoji chart. He spent 779 seconds on the first translation task, after which he felt he had scored a 50% score, a perceived result that disappointed him slightly. In reality he had scored 100%.

Anthony's translation process in the pre-setback task was very methodical. He approached each sentence analytically by determining the morphology of every word, always starting with the verb (code: *Morphological/Syntactical knowledge*). He was sequential in his actions: he showed mainly cognitive activity, which was usually followed by short bursts of metacognition if he was uncertain about something (e.g., 'no wait, that is not right.'[1]). Anthony would frequently go back and check and even recheck his assumptions.

When he observed problems (Code: *Problem defining / Evaluation*), he would laugh and try and think of alternatives. The first time he was unable to determine a case, he wrote down all the case and verb endings. He explained he did this to 'try and get a better picture in my mind'[2]. Sometimes, despite finding the correct case, he would be unsure and dither. In the end he would just go with a translation, often adding 'that is what I would do if it was a test'[3] (Code: *Just*). Just choosing a translation without being certain led him to reporting feeling uncertain when filling in the emoji chart after the pre-setback task.

After the setback, Anthony's translation process started off similarly to the pre-setback task: he began again with writing all the case and verb endings down. Again, he analysed every word grammatically (Code: *Morphological/Syntactical knowledge*). Doing so, however, his method did differ from the before the setback: he did not start with the verb but read from left to

[1] *Nee wacht, dat klopt niet.*

[2] *Ik probeer een duidelijker plaatje in mijn hoofd te krijgen.*

[3] *Dat zou ik doen als het een toets was.*

right. For the first time in the whole task, he also explicitly monitored positively 'yes this can be a ...'[1] Despite having morphologically analysed each word correctly, this was not represented in his translation, where for example the object has become subject. There was thus a discrepancy between his analysis on word level and his formulated translation of the sentence as a whole.

But, from the second sentence onwards, Anthony's translation process altered significantly. He no longer analysed the words, but he began formulating a Dutch sentence straight away using the word meanings (code: *formulating*). He eventually used some grammatical knowledge (code: *Morphological/ Syntactical knowledge*), but less extensively than before, by for example only mentioning the subject of the verb, but not the tense. Even fewer monitoring strategies were seen, and when he did, he started tapping on the desk.

Despite that the post-setback was translated much less correctly than the first set, -this time he only scored 16,6% instead of 100%- he felt he had performed better. The lack of monitoring and analysing led him to spend less time on the post-setback task: he only spent 450 seconds on it. He said that he felt most sure about this set and mentioned being motivated and hopeful after having completed the post-setback task. He thought he would score about 65%. He thus not only felt he had performed this task better than the pre-setback task, but he also felt he had performed better than he actually had done.

### TRANSLATION PROCESS UNDER EXTREME CIRCUMSTANCES: THE SETBACK TASK

In the pre-setback task, we had seen that Anthony maintained an approach in which he would analyse each word morphologically. This approach was resumed for the setback task: he wrote the case and verb endings down again and leant on cognitive knowledge. He started laughing nervously when he was dissatisfied with his translation and said for example 'I know absolutely nothing'[2] (a metacognitive observation, with a possible affective undertone). In such cases he decided to 'just write something down'[3] (code: *just*), and did not represent all the Latin words in his Dutch 'translation'.

As the task progressed, he became less and less analytical. He became distracted and restless (affective). He was frequently tapping on the desk, for example. When the researcher reminded him to think aloud, he said he was

[1] *Ja, dit kan een ... zijn.*
[2] *Ik weet helemaal niets.*
[3] *Gewoon iets opschrijven.*

disengaging[1] from the task. He then restarted by formulating a Dutch sentence based on the words he had, displaying no further metacognitive activity. By the third sentence, Anthony's approach was much quicker than in the previous two sentences. The difference in speed was caused by Anthony just stuck the word meanings together to formulate a translation without any reference to declarative knowledge. He then finished the task and stated 'I am done with it'[2]. He had spent 776 seconds on the second task. He said he felt he scored 45% because he had not completed the first two sentences and he explicitly said that left him with a bad feeling. When asked to define that bad feeling he said he was frustrated by how he had performed (affective).

## RETROSPECTIVE THINKING ALOUD

Anthony explained the frustration he felt during and immediately after the setback task in the interview: he wanted to do well but knew he had not. He mentioned that his analytical method cost him a lot of time and he took even longer to complete the tasks because he had to keep checking things again. He had not thought about this before, but questions such as 'how did you approach the first set?' and 'why did you do that?' during the interview led him to conclude that his method was awkward[3].

Anthony frequently focussed on how long he spent on the task during the interview. He mentioned having learnt in past years that he sometimes has to move on, as he otherwise would lose himself in solving the problem and not be able to finish the tasks. His way of dealing with such situations was to just write something down. This, however, frustrated him, as he wants to do it well and just moving on indicates that he is not doing well.

[1] *Ik haak een beetje af.*
[2] *Ik ben er klaar mee.*
[3] *Wat ik doe is niet echt handig bedenk ik me nu.*

APPENDIX 6.2

# NARRATIVE OF HORATIA'S TRANSLATION PROCESS

Horatia was a 15-year-old who increased her score by two on the Anagram Task after the difficult anagrams. Her self-reported mindset was incremental ($M$ = 5.3). She was only averagely affected by the difficult task, showing a decrease of two points in the post-setback task compared to the pre-setback task.

## TRANSLATION PROCESS DURING THE PRE- AND POST-SETBACK TASK

Before Horatia started the think-aloud task she reported that she felt content. She spent 517 seconds on the first translation task, after which she thought she had scored 78%. This perception was in line with the 80% she actually achieved. Horatia showed an analytical approach to translating the first set of sentences: she focused on the morpho-syntactical information of the sentence before formulating a translation. She began with determining the verb and then moved on to establishing the cases of the nouns. The information she gave was short and relevant to the translation (e.g. '*suos* is his man, plural accusative'[1]). This led to her formulating her translation in one go, with no backtracking within a sentence.

When looking at the pre-setback task, it becomes clear that Horatia relied mainly upon her cognitive knowledge of Latin, particularly knowledge of the different word endings and their meanings. After translating the first set she reported feeling uncertain on the emoji chart.

Horatia spent only 260 seconds on translating the post-setback task, far less than the 517 seconds on the pre-setback task. Her approach was generally similar to that of the first set, in that cognitive activities dominated the process. Throughout all three sentences from the post-setback task only two examples of explicit metacognitive activity were found ('wait'[2] and '-colleagues, colleague, the colleague as it is singular'[3]).

However, the cognitive activities, even though still focused on morpho-syntactical analyses, were less methodical. For example, in the first sentence Horatia mentioned nothing about the verb, whereas previously she gave more

[1] *Suos is dan zijn man, in de meervoud van accusativus.*
[2] *Wacht.*
[3] *-collega's. De collega. De collega want het is enkelvoud.*

information such as tense and person. The lack of focus on the verb led to an omission of the subject in her translation. This lack of attention for the verb was also seen elsewhere within the task, but with less consequences for the translation quality.

Particularly striking is Horatia's approach of the second sentence; she did not only use here morpho-syntactical knowledge less and less well, but she did not use it at all. The translation was made by stringing meanings of the words together to form a coherent sentence. Content-wise the sentence was correct, but the structure of the Latin sentence had not been maintained in the translation, thus leading to a lower accuracy.

When finished, the post-setback task was the task she was most certain of having translated correctly, expressing the expectation of having scored 85%. However, in reality the translations only scored 50% of the available points, a decrease in accuracy compared to her first set in which she scores 80%. However, this decrease of 2 points, was less than the average decrease.

## TRANSLATION PROCESS UNDER EXTREME CIRCUMSTANCES: THE SETBACK TASK

Horatia spent 734 seconds on the setback task. In this task we see the same emphasis on morpho-syntactical knowledge to come to a translation, that we saw her use in the pre-setback task. There were again long strings of cognitive activity and there were only two short explicit indications of metacognition, namely 'no *canorum*'[1] and 'I know that for certain'[2].

When she had completed the setback task, Horatia thought she had performed this task better compared to the pre-setback task, scoring herself 80% despite the undoable nature of the task. However, her mood by this time was still not very positive: she indicated feeling confused.

## RETROSPECTIVE THINKING ALOUD

In the interview Horatia mentioned that, despite being mostly happy with her translations, she did not understand the meaning of the sentences from the pre-setback task. She reflected on how she translated the sentences and explained

[1] *Nee, canorum.*
[2] *Dat weet ik zeker.*

that she usually ‘just about knows’[1] the translation based on the grammatical analysis she started with.

Horatia based her perceived scores on how well she had been able to translate all the words and the amount of notes she had written among the Latin texts regarding their morpho-syntax. She felt that she had translated the setback sentences well, as she did not feel like she had gotten stuck or had missed things. Another reason she gave for feeling she had done the later sets better was ‘because I had already done the first set before’[2].

She felt she had translated the post-setback sentences correctly and thought she had approached these sentences in the same way as the other sets.

[1] *Meestal weet ik het al wel een beetje.*

[2] *Maar bij sommige kon ik het ook al gewoon zien of hoefde ik er minder bij te schrijven.*

## APPENDIX 6.3

### NARRATIVE OF MARCUS' TRANSLATION PROCESS

Marcus had increased his score by four on the Anagram Task after the difficult anagrams, making his buoyancy high. His Latin proficiency was low. His self-reported mindset was non-preferential ($M$ = 3.3). His score on the post-setback task was one point higher than on the pre-setback task.

### TRANSLATION PROCESS DURING THE PRE- AND POST-SETBACK TASK

Marcus said he felt confused before starting the think-aloud task. He spent 669 seconds on the first translation task, after which he felt he had scored about half. In reality he scored 40%.

Marcus started translating the first translatable sentences by remarking 'step by step, so finite verb'[1], he thus seemed aware of using a plan (metacognition). However, Marcus was unable to work according to this plan, as he was unable find finite verbs in the Latin sentences. Marcus ended up translating by reading from left to right and looking at the word meanings. It seemed lack of certain morpho-syntactical Latin knowledge hindered him in this regard.

Generally, Marcus depended on his knowledge of the world and that of the Romans to come to translations. Many examples can be found of Marcus using general knowledge of the world or of informed guessing based on the Dutch meanings of the words. One example was that when he saw the meaning of *pilo*, he hypothesized that the sentence either meant 'the javelin pierces the horse'[2] or 'the horse is pierced by the javelin'[3]. He then discarded the second, correct, option based on the complexity of the Latin sentence: 'it does not look very complicated, so it will not be that one'[4]. The low proficient Marcus seemed to be hypothesising about the meaning of a Latin sentence and deciding between the options based on other knowledge than of Latin.

Sometimes the use of his worldly knowledge led him to realise there was a problem with his translation. In those instances, Marcus would then try to solve the problem using Latin conjugations. For example, after concluding that it seemed out of character for Caesar to keep his name out of the battle, he went

[1] *Stap voor stap, dus persoonsvorm.*
[2] *De speer doorboort het paard.*
[3] *Het paard is doorboord door de speer.*
[4] *De zin ziet er niet zo ingewikkeld uit, dus het zal niet die zijn.*

back to the Latin text and noticed the addition of -ba- to the verb. At first, such grammatical elements led to Marcus showing signs of panic ('o god'[1] and 'I hope I can persist'[2]) and he would then ramble unconnected grammatical terms whilst frequently stating that he had no idea. Marcus was, thus, able to determine that there were problems with her translation, tried to solve them and indicated that there was knowledge he might be able to use to solve them, suggesting certain procedural knowledge. However, at the same time Marcus often lacked the specific declarative knowledge to correctly solve the observed problems.

In the third set of sentences Marcus' translation process was different. He completed the third set of sentences much faster, only using 211 seconds to translate the three sentences. He was able to quickly come to translations that seemed logical to him and with one exception no mention was made throughout the task of any grammatical terms. Where in the pre-task he hypothesised regularly and gave himself options to choose from, he was content with his first option here ('I'm okay with it'[3]). This led him to move on quickly between the sentences. There was some monitoring (e.g., 'no wait'[4]) but it was sparse, particularly when compared to how he approached the first task. When Marcus finished translating, he was happy according to the Emoji Chart. This was the set he was most positive about, and he correctly evaluated that the accuracy of his translation was 50%.

## TRANSLATION PROCESS UNDER EXTREME CIRCUMSTANCES: THE SETBACK TASK

Before starting the setback task, Marcus felt curious to how he had performed the prior task. He spent 931 seconds on the second task, which was extremely long. On average the participants spent 577.63 seconds (*SD* = 215.48) on the setback task. Marcus found this task more difficult than the pre-setback task, which is reflected in the 30% score he gave himself.

The method Marcus used to try to translate the setback task was similar to how he translated the pre-setback task: he read from left to right and collected the word meanings. As he went along, he freely associated, for example after reading the word 'son' she mentioned 'Daedalus and Icarus vibes'[5]. It seems he

[1] *Oh god.*
[2] *Ik hoop dat het lukt om door te zetten.*
[3] *Ik ben er oké mee.*
[4] *Nee, wacht.*
[5] *Ik heb Daedalus en Icarus vibes.*

was trying to relate the sentences to texts he might read in Latin, in this case the word son brought the myth about Daedalus and his son Icarus to mind.

He tried to put the words into a logical order, commenting that what he had produced was strange. Again, this led him to try to conjugate. He felt that the sentence he produced 'does not sound Latiny'[1]. He then formed hypothesises about what the sentence should be and reflected on the problems those sentences had, always leaning on missing words, not on grammar. After a while he noticed that one of the ending is 'ium' which led him to exclaim 'oh god, is that not passive or something?'[2], which is not the case. Shortly afterwards he decided to give up on the first sentence as he was not able to make more of it as the word 'voice' was missing to make it logical.

Marcus, thus, continued combining the word meanings with some morpho-syntactical knowledge, whilst mainly leaning on what seemed logical on the basis of world knowledge. There were regular intervals of monitoring (metacognition) throughout the task. However, by the third sentence he was mainly negative in his assessment (e.g., 'this is totally nothing'[3]) and he finished particularly quickly as his formed 'sentence' 'at least contained all the words'. He evaluated that he certainly would only score 30% for his translation, but immediately added 'that does not matter'[4]. At this point Marcus' mood was 'interested' on the Emoji Chart and explained that he was intrigued by the sentences he had managed to produce.

## RETROSPECTIVE THINKING ALOUD

Based on his process, it seemed that Marcus was able to identify problems and knew there was knowledge he could use to solve them. However, he seemed to lack the knowledge needed. Marcus confirmed this in the retrospective interview, where he said 'I just did not conjugate properly'[5]. He seemed metacognitively aware and demonstrated good reflective skills but lacked the right cognitive knowledge to act upon his observations in a normal translation situation.

Marcus was able to explain how he should translate, but that he is unable to execute this as she lacked important knowledge. Marcus was aware of his lack of knowledge and how it limited him. He mentioned that it was a shame that his

[1] *Dit klinkt niet 'Latijnerig'.*
[2] *Oh god, is dat niet passief of zo?*
[3] *Dit is helemaal niets.*
[4] *Dat maakt niet uit.*
[5] *Ik heb gewoon niet goed vervoegd.*

knowledge of the Latin paradigms was so bad. This did not frustrate him, for him it was a given fact that he had made a choice: the number of paradigms he would need to learn to be more successful put him off of trying to learn them. Finally, he mentioned that in this instance he did not feel stress, but if it had been a test situation, he would have. In this situation he said to feel more inclined to try out different things, compared to when he is stressed, as there was less at stake.

APPENDIX 6.4

## NARRATIVE OF QUINTINA'S TRANSLATION PROCESS

Quintina was a 15-year-old whose score decreased by four on the Anagram Task after the difficult anagrams. Her self-reported mindset was non-preferential (*M* = 3.7). Her exceptionally low score in the first task (1 points) was improved upon to 3 points.

### TRANSLATION PROCESS DURING THE PRE- AND POST-SETBACK TASK

Before starting the think-aloud task Quintina felt happy. She spent only 345 seconds on the pre-setback task, after which she felt she had scored a 50% score. Slightly lower to her expectation, she scored two out of the six available points. She was among the participants who spent the least amount of time on the task.

When looking at Quintina's translation process during the pre-setback task, we saw that within the first two minutes, she mentioned finding the task difficult. Throughout the task she frequently repeated that she found it difficult (e.g. 'it is difficult'[1] or 'that is difficult'[2], seeming to confirm her low proficiency. Sometimes she commented on her lack of ability and knowledge regarding Latin (e.g. 'my grammar is not fantastic'[3]).

When Quintina tried to specifically pinpoint a problem, the problem remained generic and unconnected to Latin. Multiple examples can be found of her saying that she did not know how the Dutch words could be combined to form a sentence. In the rare cases she did try to solve the problem by looking at the Latin text, her inability to correctly use the information given in the wordlist led her to still unable to effectuate anything.

When translating, she relied heavily on the word list. Her translation strategy leant on collecting the meanings of the Latin words and creating logical sounding sentences. Often Quintina formulated two different Dutch sentences with the words and would just pick one, as she would be unsure of the best translation and during the think-aloud process she specifically told herself more than once to 'just guess[4].

---

[1] *Het is moeilijk.*
[2] *Dat is lastig.*
[3] *Mijn grammatica is niet fantastisch.*
[4] *Gewoon maar gokken.*

Quintina did explicitly demonstrate metacognition. She was not only aware that she was using guessing strategies to solve problems, but also was frequently evaluating her translations. Interestingly, she was often negative about her translation ('I do not think that it is right'[1]), but she would not try to correct something, despite knowing it was wrong.

Quintina was notable in that she spent longer on the post-setback task than on the pre-setback task, for, nearly all the participants completed the post-setback task much faster than the pre-setback task. Quintina, however spent 499 seconds on the post-setback task, over two minutes more than she spent on the pre-setback task.

In the post-setback task, Quintina's translation process was different compared to what we had seen in the pre-setback task. This time, she wrote the Dutch translations of the words above the Latin words instead of remembering them. Moreover, she frequently referred to grammatical terms and made use of case endings. Another change was that she defined problems she encountered more specifically. She, for example, said of the third sentence 'I do not know if the weapons are his'[2] and went on to explain that Octavius was the subject of the sentence and did not know if she could translate 'his weapons'.

Quintina's evaluation of her translations was more positive (e.g. I think this, yes.'[3]). This was also mirrored by her self-evaluation of her translations. This time she felt she had done better on the task but expected that there would still be mistakes. Hence, sheh gave herself a 50% score, which corresponded with the objective scores of the translation (3 out of 6 points).

## TRANSLATION PROCESS UNDER EXTREME CIRCUMSTANCES: THE SETBACK TASK

Quintina spent 562 seconds on the setback task, which was longer than on the pre-setback task. She slightly adapted her translation method compared to the post-setback task. She focused on the meaning of the individual words again, but this time, she wrote the Dutch meanings above the Latin words 'otherwise I have to keep relooking'[4]. Quintina maintained this behaviour after the setback task.

Where she had not leant on grammatical knowledge during the pre-setback task, she quickly turned to determining the cases of words in the setback task.

---

[1] *Ik denk niet dat dit klopt.*

[2] *Ik weet niet of het zijn wapens zijn.*

[3] *Ik denk dit, ja.*

[4] *Anders moet ik steeds terugkijken.*

Often the attempts were unsuccessful with her mentioning that, even if she knew which case a word was, she would not know what that meant for the translation. However, she kept trying to use it and more frequently started enumerating verb endings and even case endings correctly. In one instance, she recognised the genitive correctly, indicating certain declarative knowledge. She also knew that it was something to do with possession, but had no idea how to convey the genitive in her translation. Her knowledge still seemed to hinder her, but in a more advanced stage than earlier in the Latin Buoyancy Task. Monitoring and evaluation were also displayed to a certain extent by Quintina within the setback task. Often, she observed that something was unlikely or wrong, but her subsequent action after such an observation was usually unconnected. An example of this can be found in the first sentence 'I think dog, but it seems unlikely in this sentence. I will just go with dog'[1].

After finishing the setback task, she said that she was frustrated that she did not know how to do it and that made her feel uncertain. She felt she had performed the task slightly worse than the pre-setback task, marking her work with 40%.

## RETROSPECTIVE THINKING ALOUD

Quintina said that she had felt disheartened whilst translating because she was afraid of her results, due to finding the tasks difficult. She later mentioned that she tended to be less motivated when things were not going well and can sometimes give up. She said that the fact she was not alone whilst performing the task affected her and that she thought she would have given up on the task, particularly if she had had an answer key.

In this case she persisted however, and Quintina found that trying for longer led to trying out new and different strategies, she, for example, explained her starting to write the word meanings above the Latin: "So I thought I would just try it and it suddenly came to me that I could write it above [...] so I started doing that". As the task progressed, she began using other knowledge, such as knowledge of the world and morpho-syntactical knowledge. In the retrospective, she explained that she had forgotten about the grammar at first. She said the fact that she had decided to write the Dutch meanings of the words above the Latin meant that she was able to spend more time on other things than relooking wat the words meant. She seemed to be learning what a useful strategy was as she went.

---

[1] *Ik denk hond, maar het lijkt onwaarschijnlijk in deze zin. Ik ga gewoon hond doen.*

# APPENDIX 7.1

## WHAT STUDENTS CLAIM TO HAVE LEARNT FROM LATIN[1]

*"Vertalen is ontzettend moeilijk voor mij, ik heb geen zelfvertrouwen dus denk dan meteen dat ik het niet kan. Terwijl je bij het vertalen juist moet doorzetten om de puzzel (zin) op te lossen. Daardoor vind ik het ook niet leuk; ik ben er niet goed in."* - Roos Carlier

*"Vertalen vind ik over het algemeen niet heel leuk, want het is super moeilijk. Ik heb over mezelf geleerd dat ik niet heel goed ben in doorzetten als ik iets moeilijk vind, en hoe ik mezelf daaroverheen kan zetten. Dat is voor mij gelijk ook het belangrijkste dat ik heb geleerd, en ik kan dat ook meenemen in mijn studie en mijn verdere leven. Ik denk niet dat ik dat had kunnen leren bij een ander vak, omdat Latijn een vak is wat ik moeilijk vind, maar waarbij je niks anders kan dan doorzetten."* - Rosan Honing

*"Daarnaast heeft het vak Latijn mij echt geleerd om niet op te geven. Soms liep ik bij Latijn echt totaal vast en hoe meer dit gebeurde hoe gefrustreerder ik ervan raakte, maar toch kwam je er altijd wel uit."* - Amy van Ingen

*"Bij Latijn heb ik echt geleerd dat als ik me ergens toe zet ik hierin kan slagen. In mijn herinnering is ergens halverwege de vierde klas de knop omgegaan om me goed in te zetten voor Latijn. Voor mijn studie zal dit ook belangrijk zijn omdat dit mij vertrouwen heeft gegeven in mijn vaardigheden. Ik denk niet dat ik dit bij een ander vak had kunnen leren omdat een serieuze inzet zoals bij Latijn tot nu toe in mijn schoolcarrière niet nodig is geweest."* - Lars van Lieshout

*"Volgend jaar zou ik graag beginnen aan mijn studie Electrical Engineering aan de TU Delft, daar ga ik mijn Latijn kennis niet voor nodig hebben. Echter, zal het doorzettingsvermogen en het omgaan met tegenslagen mij hopelijk wel helpen om door te buffelen door mijn studie heen. Persoonlijk denk ik niet dat ik dit doorzettingsvermogen bij een ander vak had kunnen leren, een Latijnse tekst is een grote puzzel en heel anders dan de vakken die je simpelweg leert en kan samenvatten."* - Wessel Wagemaker

*"Aangezien ik het vertalen van Latijn best ingewikkeld vind ben ik snel geneigd om het te laten zitten. Hier heb ik de gevolgen van ondervonden en hier heb ik van geleerd dat ik, juist als ik het moeilijk vind, er meer tijd aan moet besteden en minder snel moet opgeven."* - Aukje Wiekmeijer

[1] *Original Dutch quotes*

## PROJECT OUTPUT

### PRESENTATIONS AND WORKSHOPS

O'Brien, C. (2017). *Measuring how pupils' results are affected after being confronted with undoable tasks.* Poster presentation, presented at the WCGCT World Conference, Sydney.

O'Brien, C. (2018). *Frustratie bij vertalen in de onderbouw: strategieën van leerlingen.* [Frustration when translating in the lower years: strategies employed by students] Workshop presented at the Nazomerconferentie, Nunspeet.

O'Brien, C. (2019). *Effects of an Undoable Linguistic Task on Gifted Students' Behavior.* Poster presentation, presented at the WCGCT World Conference, Nashville.

O'Brien, C. & van Rossenberg, R. J. A. (2019). HB-didactiek: leerstrategieën zichtbaar maken, [Gifted learning and instruction: learning strategies made visible] Workshop presented at GSF LIO-Themabijeenkomsten, Huizen.

O'Brien, C. (2021). *Problems Faced by Dutch Gifted Students: Examples from Latin Translation Tasks*. Presentation, presented at the online Masters Seminar 'International Perspectives on giftedness', Leipzig.

O'Brien, C. & van Rossenberg, R. J. A. (2022). *HB-didactiek*. [Gifted learning and instruction] Workshop presented at GSF LIO-Themabijeenkomsten, Huizen.

O'Brien, C. & van Rossenberg, R. J. A. (2023). *HB-didactiek: Follow-up.* [Gifted learning and instruction: Follow-up] Workshop presented at GSF LIO-Themabijeenkomsten, Huizen.

O'Brien, C. (2022). *Pushing Boundaries*. Presented at the OIKOSdagen, Ede.

# CURRICULUM VITAE

Chelsea O'Brien (1989) started teaching Latin and Greek in 2011 at the Huizermaat, a secondary school in Huizen. She has taught Classical Languages there henceforth. She completed her Master's degree in Latin and Greek Languages and Cultures at the University of Amsterdam in 2012. In 2013 she obtained her post-graduate teaching degree with honours at the Instituut voor de Lerarenopleiding (ILO) of the University of Amsterdam. In 2016, she started her part-time PhD research project at the University of Amsterdam, funded by the NWO Promotiebeurs voor Leraren. Her PhD research was supervised by Gert Rijlaarsdam and Suzanne Adema.

Besides teaching at Huizermaat, she has been a mentor of the students in the school's Gifted and Talented classes since 2012. From 2017 onwards, she has performed the task of Gifted and Talented coordinator at Huizermaat. Among other things, as coordinator, she is responsible for new admissions to the gifted classes, she provides teacher training aimed at gifted learning and instruction, she supports the gifted and talented mentors and she manages the contact between her school and other parties such as local councils and Youth Care and Protection.

Chelsea has participated in poster presentations at the World Council for Gifted and Childeren Conferences in 2017 and 2019, where she presented parts of her PhD project. In 2021, she spoke as a guest lecturer for a course on giftedness offered by the University of Leipzig. She also presented a workshop on coding and interpreting verbal transcripts of students translating Latin at a conference organized by the Association of Dutch Classicists (VCN). For the Teachers in Training programme of the Gooise Scholen Federatie, she has organized multiple workshops relating to the educational needs of gifted secondary school students.

# ACKNOWLEDGEMENTS
# (DANKWOORD)

Over the last few years, Gert has sometimes referred to this project as ‘a quest’. In literature a quest is characterized by a fixed plot, in which the protagonist pursues an important goal. Whilst trying to attain their goal, the protagonist comes across many frustrations, obstacles and temptations which endanger the successful completion of the goal. Usually, the protagonist’s journey is not so much a literal journey as a figurative journey of personal growth. Looking back, I am unable think of a better word to describe this project as than a ‘quest’. The idea of this project as a quest is further strengthened by the fact that in such stories, the protagonist is always supported by a large cast of other characters, who, each in their own way, contribute towards the end goal being achieved. Often, the supporting characters represent the brains, strength or heart. In this quest, this was no different.

Along the way, Gert and Suzanne had the somewhat dubious honour of representing all three roles. Gert, zonder jou was ik niet aan dit project begonnen, laat staan dat het afmaken was gelukt. Jouw geheel eigen en oprechte nieuwsgierigheid naar processen, jouw eerlijkheid en jouw voelsprieten maakten jou tot een ongelooflijk fijne promotor. Ik heb veel geleerd van jou over onderzoek doen en over mezelf. Toch neem ik vooral jouw manier van de vinger op de zere plek leggen en mij zo ver krijgen een stap buiten mijn comfortzone te zetten mee naar mijn begeleidingsgesprekken met mijn leerlingen. Ik wil je ontzettend bedanken voor jouw betrokkenheid als promotor.

Suzanne, als een ware Sybille die Aeneas naar de Onderwereld voert, heb je mij door dit project begeleid. Segmenteren, coderen, anagrammen maken, literatuur delen, wekelijks afspreken bij jou thuis of op allerlei locaties op de UvA, berichtjes sturen tijdens je vakanties omdat je iets zag dat je aan het onderwerp deed denken… niets was je te gek. Zelfs wanneer ik (weer) in mijn eigen valkuilen stapte, gaf jij niet op. Ik vond het heel spannend om een classica bij het project te vragen, maar jouw scherpe blik, organisatiekunsten en precisie hebben tot een flinke kwaliteitsslag geleid. Heel veel dank voor al je betrokkenheid en het meegaan op deze queeste.

*Brains*

In het kader van de ‘brains’ van deze queeste wil ik ten eerste Marcel Veenman bedanken voor het meedenken over de rol van metacognitie in het vierde hoofdstuk in het bijzonder.

Zonder alle vakdocenten die hun derde klassen hebben begeleid tijdens de verschillende fases was het onderzoek niet mogelijk geweest. Dank dat ik in jullie klassen mocht komen, dank voor het rekening houden met mijn onderzoek in jullie programma en dank voor jullie meedenken.

Nathalie, dank voor het meedenken wanneer ik vastliep op literatuur of een psychologenblik nodig had. Dank ook voor onze pogingen tot uitwisselen van onderzoeksprojecten. Juist dat we dit meermaals hebben moeten uitstellen vanwege de pubers onder onze gezamenlijke vleugels, gaf me het gevoel dat mijn project ertoe deed.

Tot slot, Bart, wat ben ik in het laatste jaar blij geweest met jouw kritische blik op strategieën (die van leerlingen of de mijne). Talent$^{HZM}$ is het begin van deze queeste geweest, dus daarom vind ik het extra bijzonder om het onderzoekstraject met jou naast me te mogen afronden. Ik wil je hier op het schakelpunt van *brains* en *strength* bedanken, omdat je mij de laatste maanden de moed hebt gegeven om door te zetten en de inhoud met mensen te delen.

*Strength*

The strength of this quest was represented by friends and colleagues. First, those from the lunchlab group. Watching and sharing in your research processes was inspiring. Each of you added to my own process, but I want to take this opportunity to specifically thank three of you. Anouk, wat heb ik het ontzettend fijn gevonden dat jij er was vanaf het allerbegin tot het eind. Marrit, jouw optimisme en oprechte interesse vond ik inspirerend. Tot slot, Daphne, ondanks al je eigen werkzaamheden was je altijd beschikbaar voor wat dan ook. Ik wil jullie dan ook hartelijk bedanken.

Veel van mijn HZM-collega's zijn erg betrokken geweest gedurende deze queeste. Een aantal wil ik in het bijzonder hier noemen. Han en Paul, dank voor jullie steun dat ik aan dit project mocht beginnen. Tanja, Valerie, Peter, Ellen en Ernee, jullie vertrouwen was er al voordat mijn beurs was toegekend. Jullie vormden gewoon spontaan een commissie waarbij ik mijn aanvraag presentatie kon oefenen! Ook al jullie vragen tijdens het onderzoekproces zelf heb ik erg gewaardeerd. Uiteraard ook dank aan al mijn lieve medementoren van de afgelopen jaren, en in het bijzonder Thomas en Fleur (met Nienke als ondersteuner). Het vertrouwen dat jullie allen hadden dat het af zou komen was soms overweldigend, maar ik hoop dat jullie weten dat jullie een belangrijk deel van het ontstaan van dit proefschrift zijn geweest. Annemieke, ik kon niet anders dan groeien onder jou als afdelingsleider. Heel veel dank daarvoor.

Noortje, Josje en Ruben, ontsnappen mogelijk maken, dat was jullie rol in deze queeste. Soms letterlijk uit een trein of een laboratorium en soms figuurlijk als ik even een ontwijkstrategie wilde inzetten. Zelfgebakken taart thuis langsbrengen tijdens lange schrijfdagen, peptalks en drankjes: jullie hebben het

allemaal gedaan! Mijn dank aan jullie is groot en ik kijk uit naar de komende ontsnappingsavonturen samen.

Jeroen, ik ben je niet vergeten hoor… maar waar moet ik beginnen? Ons co(rona)mentoraat was een avontuur op zich, maar ondertussen zag je ook mijn pieken en dalen rondom het onderzoek. Bij jou mocht ik rare statistiekvragen stellen, maar vooral ook huilen zonder tranen. Die keer dat je was blijven wachten, omdat je wist dat ik een gesprek had waar ik tegen op zag, dat was achteraf het moment dat alles weer beter werd. Je hielp me opkrabbelen, keer op keer. Heel veel dank daarvoor.

*Heart*

Een belangrijke, maar bijzondere rol in dit onderzoek namen ook al mijn (oud)leerlingen in. Zonder dat jullie dit wisten, vormden jullie het kloppende hart van dit onderzoek. Ik hoop dat mijn proefschrift jullie recht doet. Ik bedank jullie graag voor alle plezier, motivatie en inspiratie die jullie mij verlenen.

Besides my students, the heart of this project was also represented by family. Mum & dad, thank you for your confidence in me and encouraging me to push myself. Any time I needed to get away and 'just write' I was always welcomed by nan and grandad to do exactly that. Thank you so much for all the dinners and quiet evenings. Sue, you were also an integral part to that process, with your listening ear. The weeks in Barrow were particularly productive, with at least one of the chapters having been written there in its entirety. Gary, I also want to thank you for our inspiring car chats about research, writing and psychology. Those trips meant a lot to me.

Tot slot, lieve Edwin, dank je wel. Als er iemand was die van tevoren wist waarom dit hele traject misschien niet het beste idee was, was jij het. Toch liet je me deze *side-quest* niet alleen aangaan, maar ging je mee en stond je bij elke hobbel klaar om te helpen en heb je geen één keer 'zie je nou wel' gezegd. Het is nu weer tijd voor samen: samen het huishouden doen, samen koken, samen (met vrienden) kamperen, samen lang op vakantie gaan of gewoon samen even niets doen. Dank voor alles wat jij op je hebt genomen en dat je er onze queeste van hebt gemaakt. Op deze queeste was jij niet mijn Penelope of Lavinia, zoals je soms grapte, maar jij was mijn Edwin. Dank je wel.

## DISSERTATION SERIES INNOVATIVE LANGUAGE EDUCATION LAB

| | | |
|---|---|---|
| Jannemieke van de Gein (NL) | 1991 | The sense of sentences. Effects of grammar skills on aspects of writing profiency. |
| Rob Schoonen (NL) | 1991 | Constructie en validering van gestructureerde stelopdrachten voor zesdeklassers [Construction and validation of structured writing tasks for eight graders.] |
| Amos van Gelderen (NL) | 1992 | Fonetisch en linguïstische validatie van oordelen over spreekprestaties [Phonetic and linguistic validation of rating of oral performances] |
| Michel Couzijn (NL) | 1995 | Observation of Writing and Reading Activities: Effects on Learning and Transfer |
| Marianne Overmaat (NL) | 1996 | Empirische studies naar effectieve schrijfdidactiek. [Empirical Studies in Written Composition Methodology]. |
| Tanja Janssen (NL) | 1998 | Literatuuronderwijs bij benadering [Approaches in Literature Education] |
| Martine A.H. Braaksma (NL) | 2002 | Observational learning in argumentative writing. |
| Hein Broekkamp (NL) | 2003 | Task demands and test expectations. Theory and empirical research on students' preparation for a teacher-made test. |
| Isabelle de Ridder (BE) | 2003 | Hyperlinks & Vocabulary Learning from texts in L2. |
| Rita Rymenans & Sven de Maeyer (BE) | 2004 | Onderzoek naar kenmerken van effectieve scholen. Kritische factoren in een onderzoek naar schooleffectiviteit in het technisch en beroepssecundair onderwijs in Vlaanderen. [ Investigating characteristics of effective schools. Critical factors in a study of school effectiveness in technical and vocational secondary education in Flanders.] |
| Lap Quoc Trinh (Vietnam) | 2005 | Stimulating learner autonomy in English language education. A curriculum innovation study in a Vietnamese context. |
| Marleen Kieft (NL) | 2006 | The effects of adapting writing instruction to students' writing strategies. |
| Rita Rigo-Toth (Australia) | 2007 | A study of the influence of home and preschool factors on young children's literacy achievement |
| Ton Koet (NL) | 2007 | Polder English in Dutch ears. Empirical studies on the evaluation of the pronunciation of English as a foreign language. |
| Theo Witte (NL) | 2008 | Het oog van de meester. |
| Mariet Raedts (BE) | 2008 | De invloed van zelfeffectiviteitsverwachtingen, taakkennis en observerend leren bij een nieuwe en complexe schrijftaak. [The effect of self-efficacy expectations, task knowledge and observational learning in a new and complex writing task]. |
| Elke Van Steendam (BE) | 2008 | Effective instructional strategies in collaborative revision in ESL. The effects of Two empirical studies. |
| Daphne van Weijen (NL) | 2009 | Writing processes, text quality, and tasks effects Empirical studies in first and second language writing |
| Lieve Verheyden (BE) | 2010 | Achter de lijn. Vier empirische studies over ontluikende stelvaardigheid. [The story behind the line. Four empirical studies on writing by third and fourth graders of primary school. |

| | | |
|---|---|---|
| Tatiana Berden-Antonenko (NL) | 2010 | Stimulating intercultural intellectual capabilities in intercultural communication. Testing an innovative course design. |
| Talita Groenendijk (NL) | 2012 | Observe and Explore. Empirical studies on Learning in creative writing and visual arts. |
| Marion Tillema (NL) | 2012 | Writing in first and second language. Empirical studies on text quality and writing processes |
| Phuong Nam Thi Nguyen (Vietnam) | 2012 | Second language writing and literary reading in university: Three empirical studies. |
| Dick van Dijk (NL) | 2013 | De Taal van oplossingen. Een empirisch begrippenkader voor oplossingsgerichte interactie. [The Language of Solutions. An empirical conceptual framework for solution-focused interaction] |
| Milan Kríz (Tsjechië) | 2014 | Teaching Dutch as a foreign language in Eastern Europe |
| Marco Kragten (NL) | 2015 | Comprehending process diagrams in biology education |
| Marrit Hoeks-van de Guchte (NL) | 2015 | Focus on Form in Task-based Language Teaching |
| Marie-Thérèse van de Kamp (NL) | 2017 | Re-Imagine, Re-Design and Transform. Enhancing Generation and Exploration in Creative Problem Finding Processes in Visual Arts Education. |
| Marloes Schrijvers (NL) | 2019 | The story, the self, the other Developing insight into human nature in the literature classroom. |
| Klaske Elving (NL) | 2019 | Effectieve Leeractiviteiten voor het Schrijfonderwijs in HAVO 4 [Effective Learning Activities for the Teaching of Writing in General secondary education, grade 10]. |
| Suzanne Luger (NL) | 2020 | Lost in Latin Translation. Teaching students to produce coherent target texts |
| Nina Vandermeulen (BE) | 2020 | Synthesis Writing in upper-secondary Education |
| Magdalena Flores-Ferrés (Chile) | 2021 | Teaching Writing in Chile. An evidence base for policies on writing instruction for Grades 7-12 of Chilean public schools |
| Martijn Koek (NL) | 2022 | Think Twice: Literature Lesson that Matter. |
| Liselore van Ockenburg (NL) | 2023 | Synthesis writing. Teaching high school students how to read, plan, draft, and revise |
| Edith Alkema (NL) | 2023 | Schrijven met kennis van zaken. Een didactiek voor de zaakvakken. [Knowledgeable writing. Writing instuction for social sciences] |