Undergraduate students' performance in mathematics: Individual and combined effects of approaches to learning, self-efficacy, and prior mathematics knowledge

Yusuf F. Zakariya

UiA University of Agder

Doctoral Dissertations at the University of Agder 312

Yusuf F. Zakariya

Undergraduate students’ performance in mathematics: Individual and combined effects of approaches to learning, self-efficacy, and prior mathematics knowledge

Dissertation for the degree of philosophiae doctor

University of Agder
Faculty of Engineering Science
2021

Doctoral dissertations at the University of Agder 312

ISSN: 1504-9272
ISBN: 978-82-8427-016-6

Print: 07 Media
Kristiansand, Norway

Acknowledgements

All praises and adorations belong to Almighty God for the wisdom He bestowed on me through the process of the present study. I acknowledge the effort of my supervisor team for their continued support for the successful research. Many thanks to members of my family for the patience and perseverance during the time Daddy is not at home. The teaching and the non-teaching staff of the Department of Mathematical Sciences at the University of Agder, Faculty of Engineering administrators, and the university library are appreciated for their solidarity in creating a conducive environment for conducting the present research.

Summary

This dissertation concerns an exploration of factors that affect the learning outcomes of students in higher education mathematics. It is framed within a quantitative research paradigm in which the existence of personal factors such as prior knowledge, self-efficacy, and approaches to learning are assumed and that these factors may be operationalised and measured. The aim of this study is to investigate effects of prior mathematics knowledge, approaches to learning and self-efficacy on students' performance in a first-year introductory calculus course for students on engineering programmes. Further, the interrelatedness of these factors and their combined effects on performance are also investigated. Two well-established psychological theories are combined to form the conceptual framework for justifying appropriateness and usefulness of chosen constructs under investigation coupled with hypothesised relationships among the constructs. These theories are student approaches to learning theory and self-efficacy theory. A cross-sectional survey research design was adopted with a focus on engineering students aimed at addressing three research questions. These research questions are formulated as follow:

1. Do approaches to learning mathematics differ with respect to the prevalence of deep and surface approaches among first-year engineering students?
2. Does self-efficacy influence adoption of either deep or surface approach to learning mathematics among first-year engineering students?
3. What are direct and indirect effects of prior mathematics knowledge, approaches to learning, and self-efficacy on performance in mathematics among first-year engineering students?

The data used for the present study were collected in two phases (pilot study and main study) using questionnaires, a pre-test of students' basic mathematical knowledge and final examination scores in an introductory calculus course. The pilot study data were collected in Spring 2019 and used to develop and validate the questionnaires. The main study data were collected in Autumn 2019 and used to investigate hypothesised structural relationships between prior mathematics knowledge, approaches to learning, self-efficacy, and students' performance in the course. Eight research hypotheses were formulated and tested using structural equation modelling techniques. The resulting findings were well-documented and published in seven peer-reviewed journal papers and one peer-reviewed conference paper.

Paper I and Paper II present results of validity studies on a measure of approaches to learning mathematics. Therein, psychometric properties such as construct validity and reliability of two-factor revised study process questionnaire (R-SPQ-2F, Norwegian version) are reported. Further, the findings reported in Paper I and Paper II confirm the prevalence of deep and surface approaches to learning mathematics among the first-year engineering students. Thus, the findings address the research question one. A measure of self-efficacy of students on calculus tasks was developed and validated in Paper III. Therein, self-efficacy was conceptualised and operationalised based on postulates of the self-efficacy theory. The findings confirm an acceptable construct validity, discriminant validity and reliability of the measure. Paper IV and Paper V (conference paper) present results of validity studies on a test of prior mathematics knowledge. Therein, item quality such as difficulty indices, discrimination indices, and item reliability were studied using item response theory. The test was revised based on the findings of the studies reported in Paper IV and Paper V. Some items were removed from the test before being used in the analyses of the main study data.

The research question two was addressed in Paper VI. Therein, the results show that self-efficacy has a positive causal effect on deep approaches to learning and a negative causal effect on surface approaches to learning. Thus, I argue that engineering students' approaches to learning mathematics may be influenced by fostering their self-efficacy through interventions. The research question three was split into two questions and addressed in Paper VII and Paper VIII. Therein, I observed that prior mathematics knowledge test has substantial negative and positive effects on surface approaches to learning and self-efficacy, respectively. However, its effect on performance was only significant when I screened out self-efficacy from the structural model. The surface approaches to learning have a negative effect on students' performance in the course. In contrast, there was no substantial evidence to justify any considerable effect of the deep approaches to learning on students' performance.

More so, the results show that self-efficacy has a substantial positive causal effect on students' performance in the calculus course. The findings further reveal that both surface approaches to learning and self-efficacy play a mediating role between prior mathematics knowledge and performance. Finally, I argue that since self-efficacy influences adoption of approaches to learning, it is prudent to develop interventions that foster self-efficacy as proxies to influence both approaches to learning and students' performance, simultaneously. As such, I make some recommendations for future studies on possible interventions that foster self-efficacy.

The resulting findings from the eight papers knitted together in the present dissertation make crucial contributions to literature at the local, national, and global levels. First, because of the cultural sensitivity of R-SPQ-2F coupled with its lack of psychometric study within the Norwegian context, I claim that a validation of the measure is a crucial contribution to the literature. Second, through the present study, the Norwegian Mathematical Council test was validated for the first time, over the past three decades. The findings and recommendations for the test improvement were communicated to the Council, which I perceived to be a crucial contribution at the national level. Further, the new measure of calculus self-efficacy that was developed through this study constitutes an original contribution to literature. The reason being that extensive search of the literature reveals such a specific measure of self-efficacy is unprecedented. Other sets of contributions of the present study to the literature emanate from the item-level structural equation modelling used in evaluating the hypothesised relationships between the research constructs. This modelling technique is a paradigm shift from the multiple linear regression studies that are commonly reported in mathematics education literature. Thus, I believe that university teachers, researchers, policymakers, and other stakeholders involved in teaching first-year undergraduate mathematics courses will benefit optimally from the findings reported in the present dissertation.

Table of contents

List of Figures

List of Tables

List of papers

1 Introduction

1.1 Background

The poor performance of students in undergraduate mathematics has gained increased global attention among education researchers (Eklund, 2019; Eng, Li, & Julaihi, 2010) in which Norway is not an exception. Gynnild, Tyssedal, and Lorentzen's (2005) report of 21.5% to 39.2% failure rates over five years in a first-year introductory calculus course at a university in Norway is a typical example. More so, results of a descriptive analysis on grade distributions in an introductory calculus course among first-year engineering students at a Norwegian university show that the problem of poor performance in mathematics persists within the university. I examined these students' grades for Autumn 2019 semester and found that only six students (2%) got As, 18 students got Bs, 51 students got Cs, 43 students got Ds, 56 students got Es, and 130 (43%) students failed the course.

This prevalence of poor performance among engineering students has been worrisome to mathematics teachers, education researchers, Centre for Research, Innovation and Coordination of Mathematics Teaching (MatRIC), and other stakeholders in teaching and learning of mathematics at the university. Thus, the rationale behind the conduct of the present study is to find possible solutions to the problem of poor performance in mathematics with a focus on first-year engineering students. I focus on first-year engineering students because they are more susceptible to poor performance in mathematics as exposed in the grade distributions that I presented in the previous paragraph. The question is how to find these solutions? This question constituted my first puzzle when I accepted PhD fellowship at the University of Agder.

Diverse studies abound in the literature that provides empirical evidence for various factors that affect students' performance in mathematics. These factors are from different sources of influence. These sources can be home, e.g., socio-economic status (Wang, Li, & Li, 2014), school, e.g., leadership of the school (Tan, 2018), classroom structures, e.g., class size (Konstantopoulos & Shen, 2016), curricula, e.g., use of calculators (Mao, White, Sadler, & Sonnert, 2016), students, e.g., approaches to learning (Maciejewski & Merchant, 2016), and teachers, e.g., teachers' attributes and teaching methods (Zengin, 2017).

The present study (being a pioneering project of its kinds in terms of the adopted quantitative research methodology at the university) has a focus on those factors that emanate from the students (i.e., student-source factors) and that have a strong influence on performance in a first-year mathematics course. An underlying assumption governing the decision of focusing on student-source factors is that these factors can be influenced and improved upon by providing interventions and remedial action while working with the students. Another challenge for me at the initial stage of the present study was in the identification of manageable student-source factors within the timeline of the PhD programme. In order to overcome this challenge, I embarked on a literature review which is summarised in the next section. A comprehensive exposition of the reviewed literature is available in the published papers that are included later in this dissertation.

1.2 Student-source factors and performance in mathematics

Researchers across the globe have studied the relationships between factors with sources from the students, i.e., students' personal characteristics, and performance in mathematics. It is not my intention to provide an exhaustive review of these studies. However, some prominent studies on these factors are worth mentioning because of their relevance to the research reported herein. These factors include students' self-concept beliefs in mathematics (e.g., Pajares & Miller, 1994), their mathematics motivation (e.g., Tossavainen, Rensaa, & Johansson, 2019), mathematics conceptions (e.g., Yang, Leung, & Zhang, 2019), students' learning approaches (e.g., Maciejewski & Merchant, 2016), prior knowledge of mathematics (e.g., Rach & Ufer, 2020), self-efficacy (e.g., Williams & Williams, 2010), attitude towards mathematics (e.g., Dowker, Cheriton, Horton, & Mark, 2019), mental ability (e.g., Pajares & Kranzler, 1995), and students' anxiety about mathematics (e.g., Dowker, Sarkar, & Looi, 2016). However, the prior mathematics knowledge, self-efficacy and approaches to learning stand out in terms of their strong influence on performance in mathematics and their better predictions of performance than some other student-source factors as it will be highlighted in the forthcoming sections. As such, I focus on these three factors in the present study.

1.2.1 Approaches to learning

Approaches to learning have been conceptualised to encapsulate predispositions of an individual when presented with learning materials and the strategies used to process the learning contents (Biggs & Tang, 2007). Students adopt various approaches when presented with learning tasks. However, it has been theoretically established that these various approaches may be sufficiently characterised as deep and surface approaches (Marton & Booth, 1997). Deep approaches to learning connote the processes of high cognitive activity in which students concentrate on developing a proper understanding of learning materials. In contrast, surface approaches to learning are processes of low cognitive activity in which students focus on passing the course while doing minimal work as possible (Biggs, Kember, & Leung, 2001; Entwistle, 2005). As such, students who adopt surface approaches to learning usually resort to memorisation of key concepts and techniques in the learning materials (Marton & Säljö, 2005). Social-psychological theorists (e.g., Marton & Säljö, 2005) have argued and shown empirically that approaches to learning are context-specific and grossly influenced by students' intention. Therefore, approaches to learning in the present study are adopted processes by or predispositions of engineering students toward learning a first-year introductory mathematics course.

Education researchers have established strong relationships between approaches to learning and the nature of presented mathematics tasks, mathematics conceptions, and attitude towards mathematics (Alkhateeb & Hammoudi, 2006; Maciejewski & Merchant, 2016; Mji, 2000). It has also been shown empirically that approaches to learning are better predictors of students' performance in mathematics than their level of mathematics anxiety, gender, mathematics motivation, the utility of mathematics, and the enjoyment of mathematics (García, Rodríguez, Betts, Areces, & González-Castro, 2016). As a result, I presume that it is prudent to focus on approaches to learning in the present study instead of the mathematics anxiety, gender, mathematics motivation, the utility of mathematics, and the enjoyment of mathematics.

However, the findings concerning the specific influence of either deep or surface approaches to learning on performance in mathematics are inconsistent. In some studies (e.g., Cano, Martin, Ginns, & Berbén, 2018), deep approaches to learning predict students' performance in mathematics while surface approaches do not.

Some researchers (e.g., Nguyen, 2016) reported the opposite, i.e., surface approaches to learning predict performance in mathematics while deep approaches to learning do not. At the other extreme, Mundia and Metussin (2019) found that both deep and surface approaches have no substantial influence on students' performance in mathematics. These contrasting findings expose a gap in mathematics education literature in which the present study intends to address.

1.2.2 Self-efficacy

Studies on learning experience in higher education are not limited to the students' adopted approaches to learning. A good number of psychologists and sociologists have dug deep into students' reflections of themselves as they learn (Bandura, 1986; Hackett & Betz, 1989; Pajares, 1996). An outcome of this deep insight into students' learning is the identification of perceived self-efficacy as a strong influencer of desirable learning outcomes (Bandura, 1993). Perceived self-efficacy is conceptualised as the "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). It has been demonstrated to be a crucial student personal factor that enhances perseverance when encountering difficult learning tasks and a drive towards the attainment of high achievement during the learning activity. Decades of both theoretical and analytical studies on perceived self-efficacy have confirmed its task-specificity as it concerns mathematics learning (e.g., Pajares & Miller, 1995). As such, a working definition of perceived self-efficacy in the present study encompasses students' convictions about their competence to solve first-year mathematics tasks successfully.

Previous studies have confirmed that students with a high sense of perceived self-efficacy are highly motivated to learn, develop positive attitudes toward mathematics, are highly interested in mathematics, and have low mathematics anxiety (Bandura, 1997). Perceived self-efficacy predicts students' performance on mathematics problem-solving activities better than the mental ability, mathematics anxiety, self-concept, the utility of mathematics and prior mathematics knowledge (Pajares & Kranzler, 1995; Pajares & Miller, 1994). As such, I presume that it is prudent to focus on perceived self-efficacy in the present study instead of the mental ability, and mathematics self-concept.

Despite the advantage of perceived self-efficacy over other student-source factors, in terms of its better predictive power of performance in mathematics, limited studies are available on its causal effects on students' performance in university mathematics. I believe that for perceived self-efficacy interventions to be effective as proxies for improved performance in mathematics, one must establish a causal relationship between the two variables. Thus, the present study attempts to fill this gap.

1.2.3 Prior mathematics knowledge

There are different levels (in terms of course contents and difficulty) of secondary school mathematics that engineering students followed before their enrolment into university programmes in Norway. These diverse mathematics courses make it difficult to find a coherent common descriptor of prior mathematics knowledge that is suitable for the Norwegian context. As such, it may be unrealistic to meet all the requisite elements that are embedded in the proposed conceptualisation of prior knowledge by Dochy, De Rijdt, and Dyck (2002) as follows:

> The whole of a person's knowledge, which is as such dynamic in nature, is available before a certain learning task, is structured, can exist in multiple states (i.e. declarative, procedural and conditional knowledge), is both explicit and tacit in nature and contains conceptual and metacognitive knowledge components (p. 267).

As an interim, I conceptualise prior mathematics knowledge as students' performance on a test that is designed to expose Norwegian secondary school basic mathematics content knowledge. Prior mathematics knowledge is a crucial factor that affects the current mathematics learning outcomes. Students who are well prepared, in terms of the requisite mathematics content knowledge, for university learning clearly have a better chance of succeeding in the first-year mathematics courses than those students who are ill-prepared. Prior mathematics knowledge has been theoretically argued to be a potential source of perceived self-efficacy (Bandura, 1997; Usher & Pajares, 2009). Further, there is an accumulation of evidence on the direct influence of prior mathematics knowledge on approaches to learning and the students' current performance on mathematics tasks (e.g., Hailikari, Nevgi, & Komulainen, 2008; Nguyen, 2016). However, little is known about the indirect effects of prior mathematics knowledge through either students' approaches to learning or perceived self-efficacy on performance in mathematics. Therefore, it is part of the intention of embarking on the present study to address this gap.

1.3 Knowledge gaps and the research aims

The summary of previous studies on the relationships between student-source factors and performance in mathematics that is presented in Section 1.2 exposes some knowledge gaps in mathematics education literature which the present study attempts to fill. The relationships between prior mathematics knowledge, the approaches to learning the subject, and perceived self-efficacy coupled with their combined effects on performance in mathematics have been sparsely reported in the literature. The available studies lack coherence in their findings. Some are conducted using statistical approaches that are not appropriate for evaluating causal claims. While others are conducted outside university mathematics learning whose findings are not generalisable to the population of the present study due to the task-specificity of the constructs under investigation. Further, to the best of my knowledge, after an extensive search of the literature, the interplay between these factors and their indirect effects on performance in university mathematics have not been reported in the Norwegian context.

In a bid to fill these knowledge gaps, the present study is formulated to investigate causal effects[1] of prior mathematics knowledge, approaches to learning mathematics, and perceived self-efficacy on students' performance in a first-year introductory calculus course. Further, the interrelatedness of these factors and their combined effects on performance in mathematics are also investigated. I believe that my attempts in achieving these aims will serve crucial purposes in informing efforts to alleviate the problem of poor performance in mathematics among first-year engineering students by identifying factor(s) to be prioritised for interventions. Further, the findings of the present study will be beneficial to university teachers, education researchers, policymakers, and other stakeholders

---

[1] The causal effect that is intended at this point is a functional relationship between the research variables such that changes in a causal variable lead to changes in the probability distributions of the effect variables. The definition may be contrasted with deterministic causation that requires that a change in causal variable leads to the same change at all levels of the effect variable (Kline, 2016).

who are involved in teaching and learning first-year undergraduate mathematics courses for improved performance.

1.4 Research questions

To achieve the research aims that are presented in Section 1.3, I formulate and attempt to address the following research questions in the present study:

1. Do approaches to learning mathematics differ with respect to the prevalence of deep and surface approaches among first-year engineering students?
2. Does perceived self-efficacy influence adoption of either deep or surface approach to learning mathematics among first-year engineering students?
3. What are direct and indirect causal effects of prior mathematics knowledge, approaches to learning, and perceived self-efficacy on performance in mathematics among first-year engineering students?

1.5 Outline of the PhD study

The three research questions presented in Section 1.4 are the central queries the present study attempts to address. As such, I followed a two-stage approach to scientific inquiry in the present study. A pilot study stage and a main study stage. The pilot study was aimed at the development and the validation of measures of the research constructs. The main study focused on the structural validation of hypothesised relationships between the research constructs. Structural equation modelling approach was mainly used in the analyses of the collected data for both the pilot and the main studies. The structural equation modelling approach pens the opportunity to evaluate causal claims as well as to deduce *causal relationships*[2] between the research variables. These causal claims are either inappropriate or rather unrealistic to be evaluated using other competing models such as the classical multiple linear regression, analysis of (co)variance, and path analysis (Bollen & Pearl, 2013).

---

[2] There is a causal relationship from a variable P to a variable Q if

a. P has a non-trivial correlation with Q,
b. P temporary precedes Q,
c. and there is no variable R with confounding effects on both P and Q (Antonakis, Bendahan, Jacquart, & Lalive, 2010).

Further, if such variable R exists, which is always the case, it should be controlled for in the structural model.

I believe that if I can provide empirical evidence that:

a. characterises engineering students' approaches to learning mathematics
b. establishes the contribution of prior mathematics knowledge to engineering students' performance in mathematics, and to what extent this contribution influences?
c. exposes the main students' approaches to learning (either deep or surface) that influence performance in mathematics, and to what extent?
d. establishes the contribution of perceived self-efficacy to the students' performance in mathematics, and to what extent?

Then, by implication, I can possibly point to where interventions might have a positive impact and suggest (describe) evidence-based interventions to help engineering students perform very well in introductory mathematics courses.

1.6 Outline of the present dissertation

The remaining parts of the present dissertation are arranged in chapters. The next chapter that follows the current introduction chapter focuses on the conceptual framework. Therein, I introduce different types of frameworks in mathematics and argued for my choice of using the conceptual framework. Then, I present the primary postulations of both student approaches to learning theory and the self-efficacy theory and link these postulations to my hypothesis formulations. The chapter concludes with some strengths, limitations, and my reflections on potential ways to network the theories for a coherence argument in the present study.

Chapter Three is on methodology and methods. In the chapter, I highlight the crucial elements of my research paradigm and my conception of measurement. I introduce the specifics of operationalisations and measures of the research constructs. Further, I delve into some salient issues on validity and reliability as they relate to the latent variable theory of measurement. I immediately follow these issues with some procedures for data collection and analyses. The chapter concludes by highlighting some important aspects of ethical considerations in the present study.

I present an overview of the papers in Chapter Four. For each paper, I highlight the study aims, specific research methods, and primary findings. Further, I demonstrate how the findings of each paper address the research questions and or hypotheses they are purported to address. Chapter Five presents an elaborate discussion on the findings that are highlighted in Chapter Four. I expose some implications of these findings for agents of implementation, e.g., mathematics teachers and curriculum planners. I register my dispositions toward the concept of test validity, the intended validity evidence exposed in papers I-V, and the necessity for more validation studies to strengthen my findings. Further, I reflect on the contributions of each paper toward making a coherence argument for the achievement of the research aims. Finally, I acknowledge some potential limitations of this study and the implications of these limitations to generalisation of findings.

In Chapter Six, the present dissertation concludes by enumerating the steps taken through the course of the present study to communicate the findings to the agents of implementation. I argue for the significance of the present study and its unique contributions to the literature. Finally, I highlight the most significant result of the present study and some descriptions of evidence-based interventions as proxies to enhance engineering students' performance in an introductory calculus course.

2 The conceptual framework

2.1 Frameworks in mathematics education research

The online Oxford advanced learner dictionary defines a framework as “a set of beliefs, ideas or rules that is used as the basis for making judgements, decisions, etc.” In the context of education research, a framework has been described as “a basic structure of the ideas (i.e., abstractions and relationships) that serve as the basis for a phenomenon that is to be investigated” (Lester, 2010, p. 69). Thus, adopting a framework in mathematics education research comes with lots of advantages. Some of these advantages are provisions of structures for conceptualisations of research constructs; formulating research questions; hypothesising relationships between research constructs; selecting or developing research instruments; justifying research methods; making sense of research data; and interpretations of results (Eisenhart, 1991; Lester, 2010). These advantages manifest through the crucial roles of theories in such frameworks. Theories in mathematics education serve many purposes such as lenses to examine the data, tools for analysing pedagogical activities, descriptions of the essence of learning, and arguments for justifying relationships between research constructs (Lester, 2010; Prediger, Bikner-Ahsbahs, & Arzarello, 2008).

In a paper presented at the thirteenth annual meeting of the North American chapter of the international group for Psychology of Mathematics Education, Eisenhart (1991) distinguished between three types of education research frameworks: (a) *conceptual framework*, (b) *theoretical framework*, and (c) *practical framework*. The main difference between these research frameworks boils down to the role of theory or theories therein. In a theoretical framework, for example, a theory (or theories) is usually assumed as the framework itself such that all the research activities are guided and explained through the adopted theory. For example, Jean Piaget’s research is deeply rooted in a constructivist genetic epistemology of adaptation (assimilation and accommodation). At the other extreme, is the practical framework (early work of Michael Scriven is an example (e.g.,Scriven, 1986)), in which case, there is little or no recourse to a theory. The research activities are based on ‘what works’ and the experience of the researchers (Lester, 2010).

At the middle of these two frameworks (theoretical and practical) lies the conceptual framework. A conceptual framework according to Eisenhart (1991) is "a skeletal structure of justification, rather than a skeletal structure of explanation based on logic (i.e., formal theory) or accumulated experience (i.e. practitioner knowledge)", (p. 209). One distinguishing property of the conceptual framework is that a theory (or a variety of theories) is used as arguments for justifying appropriateness and usefulness of chosen constructs under investigation coupled with any expected relationship among them. This justification happens without total submission to every dictate of the proponents of such theories. This view is contrary to the theoretical framework in which researchers mostly see a theory (or theories) as a lens (perhaps the only one) through which data are viewed such that hypotheses of the theory are tested with the aim of supporting, modifying, or extending the theory. More so, the use of the theoretical framework in mathematics education research has been partly criticised for enforcing interpretation of results based on a 'theoretical decree' rather than evidence (i.e. what the data say?), lack of 'triangulation', and encouraging localised meanings (Lester, 2010).

It is my opinion that adopting any of the three types of frameworks in mathematics education research depends on several factors such as the research focus, the research paradigm, and the researcher's inquiry aims, rather than a mere right or wrong dichotomy. Thus, I adopt a conceptual framework in the present study for two reasons. First, I consider the nature of my study, which is framed within the quantitative research paradigm[3]. Second, I consider my inquiry aims which are to investigate potential causal relationships between the research constructs, unlike a description of the essence of such constructs (typical of a qualitative research paradigm). As such, I erect the skeletal structure for the present study by combining ideas from two well-established psychological theories and previous studies to operationalise and justify hypothesised relationships between my research constructs. The operationalisation of my object of research is achieved through the use of the theories and previous studies in formulating my research hypotheses as it is typical of research within the quantitative paradigm (e.g., Maciejewski & Merchant, 2016; Schukajlow, Achmetli, & Rakoczy, 2019). These two theories are student approaches to learning (SAL) theory (Marton & Säljö, 1976a, 2005) and self-efficacy theory (Bandura, 1977, 2012).

[3] An elaborate discussion on the research paradigm is presented in the next chapter.

2.2 Student approaches to learning (SAL) theory

2.2.1 Origin, ontology, and epistemology of the SAL theory

Marton and Säljö dissatisfied with the dominant information processing (IP) theory in their research group at that time, went ahead to develop the SAL theory from their series of qualitative experimental studies. These studies were focused on Swedish undergraduate students' approaches to reading, understanding and answering questions based on some presented passages of prose and newspaper articles (Marton & Säljö, 1976b, 2005). The ontology of this theory is perceived to be critical realism in which there is a rejection of multiple realities for the world people live in. To the SAL tradition, an individual learns (i.e. gains knowledge about the world through experiencing) in a dialectic relationship with social factors and that the knowledge gained is not constructed individually nor imposed by the environment (Marton & Booth, 1997). Therefore, individual's characteristics, such as approaches to learning, are being shaped dynamically by social factors, prior experience and conception of learning (Marton & Booth, 1997). Thus, reality (in this case, learning) according to SAL theory is situated within the teaching-learning context as opposed to the IP theory that restricts learning conception to what happens within an individual (Marton & Booth, 1997). Even though SAL theory originates from studies on newspaper passages it has been modified, expanded and applied to diverse teaching and learning situations (e.g., Hounsell, 2005).

In the original experiments conducted by Marton and Säljö, they utilised the term "approaches to learning" to connote adopted *processes* by the students prior to the experiments which directly affect their learning outcomes. The epistemology of the SAL theory as pursed by Marton and Säljö is *phenomenography.* The phenomenography is "a way of identifying, formulating, and tackling certain sorts of research questions, a specialization that is particularly aimed at questions of relevance to learning and understanding in an educational setting" (Marton & Booth, 1997, p. 111). It relies on students' exploration of experience, prior knowledge and students' description of what learning means to them. It encompasses a bottom-up qualitative research methodology in which a researcher explores a phenomenon without a formal theory before data collection. Instead, the evidence is extracted from the students' account of the learning context to establish a coherent argument for describing the phenomenon (Marton, 1981).

2.2.2 Approaches to learning are not fixed constructs

A basic tenet of SAL theory is that students approach learning in different ways because of constant interactions between intention, motive, and learning context. However, these diverse approaches can be classified into finite, manageable categories; in fact, two categories – deep and surface approaches (Marton & Säljö, 2005). This historic categorisation of students' learning approaches has equally been confirmed in other studies (e.g., Biggs, 1987; Entwistle & Waterston, 1988; Svensson, 2005). The SAL theory is in contrast with the IP theory which sees students' learning approaches as a function of mainly cognitive ability, fixed traits and universal across all cultures (Moreno & DiVesta, 1991). The IP theoretical framework has received criticisms for its perceived inappropriateness to describe students' learning approaches as it excludes a good number of social factors. Some of these factors were enumerated by Biggs et al. (2001) as "students' values and motives, their perceptions of task demands, teaching and assessment methods, classroom climate, and so on." (p. 134). Thus, it follows that approaches to learning mathematics (deep and surface) among engineering students are not fixed but change as learning situations change coupled with students' affective reactions.

Students that adopt deep approaches to learning get hold of information with the intent of discovering the intended meaning of the learning material. In contrast, surface approach learners are preoccupied with the discourse or the text itself with little or no attention to the intended meanings. In more succinct words, "the former refers to paying attention to the meaning and significance of the materials to be learned, whereas the latter concentrates more on rote memorising" (Lonka, Olkinuora, & Mäkinen, 2004, p. 302). More recently, Biggs (2012) while distinguishing between the surface and deep approaches to learning posited that the surface approach to learning "refers to activities of an inappropriately low cognitive level, which yields fragmented outcomes that do not convey the meaning of the encounter" and the deep approach to learning "refers to activities that are appropriate to handling the task so that an appropriate outcome is achieved.", (p. 42). As such, I argue that because engineering students are being trained to solve practical problems using mathematics content knowledge, their approaches to learning introductory calculus course are expected to be deep approaches. However, since I cannot rule out students who are interested only in achieving a passing or good grade on the course, surface approaches to learning are also plausible among the engineering students.

Therefore, the following hypothesis is formulated:

> *Hypothesis 1*: There are differences in calculus learning approaches among first-year engineering students in terms of the prevalence of deep and surface approaches.

2.2.3 Conception of mathematics, motivation, and approaches to learning

The SAL theory posits that the type of approaches (deep or surface) adopted by students in learning a content is predictable from their conceptions of learning which in turn is linked with motivation to learn (Marton & Säljö, 2005). With a focus on mathematics learning, students who conceive the calculus course as something useful and proper understanding of it is necessary for intellectual development tend to adopt deep approaches to learning the course. In contrast, students whose conception of the calculus course is just a requirement to move to the next level of study are more likely to adopt surface approaches to learning the course. To substantiate this claim, Mji (2003) found that there is a strong relationship between surface approaches to learning and students' conceptions of mathematics in a longitudinal study that involves 459 undergraduate students enrolled on a first-year mathematics course. Thus, deep approaches to learning are intrinsically motivated by the intention to develop conceptual understanding of learning materials, while surface approaches to learning are extrinsically motivated by the quest to achieve good grades in the course (Hounsell, 2005; Marton & Säljö, 2005).

2.2.4 Relationships between approaches to learning, previous and current performance

Following Biggs et al. (2001) lines of thought, I deem it necessary to remark that SAL theory cannot be disjointed from the presage-process-product (3P) model initially developed by Dunkin and Biddle (1974) and adapted from classroom teaching context to student learning context by Biggs (1993). The 3P model elicits a dynamic interaction between *presage factors* (e.g., prior knowledge), *process factors* (e.g., ongoing approaches to learning) and *product factors* (e.g., students' learning outcomes), as shown in Figure 1. This framework has helped educational psychologists in interpreting SAL theory and in developing instruments for measuring the approaches to learning (Biggs et al., 2001; Entwistle & Tait, 1994). In doing so, a more refined conceptualisation of *approaches to learning* as motives, predispositions, styles, and strategies to adopt a process was proposed (Biggs, 1993).

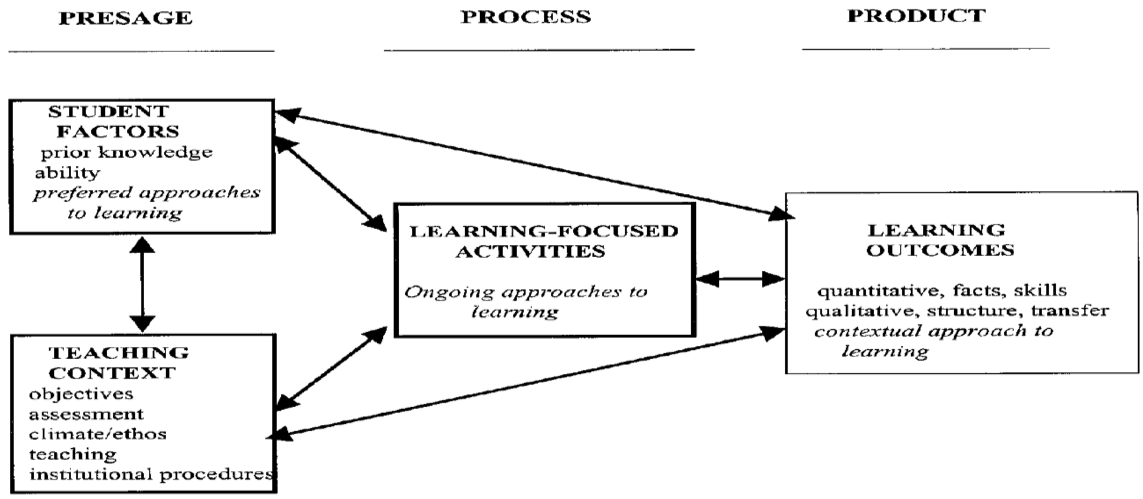


Figure 1. The 3P model of teaching and learning

Note. Reprinted from "The revised two-factor Study Process Questionnaire: R-SPQ-2F" by J. Biggs, D. Kember and D.Y.P. Leung, 2001, *British Journal of Educational Psychology*, *71*(1), p. 136 (https://doi.org/10.1348/000709901158433). Copyright 2001 by the British Psychological Society.

Figure 1 shows a dynamic system of continuous interaction between the presage, process, and product factors with ongoing approaches to learning at the centre of the model. The double-headed arrows indicate the feedback relationships between components of this model. Therefore, it can be deduced from the 3P model that engineering students' learning approaches are perceived as context-dependent that change from one context to another based on prior mathematics knowledge and students' current performance in mathematics.

SAL theory has gained wide acceptance among education researchers in diverse fields, e.g., mathematics, science, and engineering especially in developing its measures and predicting students' learning outcomes (e.g., Asikainen & Gijbels, 2017; Biggs et al., 2001; Biggs & Tang, 2007). Deep approaches to learning are generally associated with increased learning outcomes, while surface approaches to learning are generally associated with decreased learning outcomes. For instance, Maciejewski and Merchant (2016) found in their study that deep approaches to learning have a positive correlation with student mathematics grades in the first year, while surface approaches to learning have a negative correlation with student mathematics grades in year two, year three and year four. Moreover, relying on SAL theory directly or on measures based on SAL theory, some education researchers have investigated approaches to learning from a domain-specific perspective such as mathematics, civil engineering, and economics (e.g., Maciejewski & Merchant, 2016; Salmisto, Postareff, & Nokelainen, 2017).

Therefore, based on the postulates of SAL theory coupled with some insights from the previous studies, I formulated the following hypotheses:

> *Hypothesis 2*: There is an effect of prior mathematics knowledge on engineering students' ongoing approaches to learning.
> *Hypothesis 3*: There are effects of engineering students' approaches to learning on their performance in a first-year calculus course.
> *Hypothesis 4*: Ongoing approaches to learning mediate the effect of engineering students' prior mathematics knowledge on their performance in a first-year calculus course.

It is important to remark that the 3P model in the SAL tradition postulates a dynamic feedback relationship between approaches to learning, prior knowledge, and current students' performance. However, hypotheses 2 to 4 present one-directional effects between these constructs. It is my opinion that a dynamic feedback relationship is best investigated especially for these constructs using a longitudinal research design. Meanwhile, the present study follows a cross-sectional design due to some external constraints of the research. It will be interesting to conduct a future study with this intention. More so, hypotheses 2 to 4 have been formulated, in a broad sense, without differentiations into specific signs (positive or negative) of effects of each construct on another because these signs of effects are part of the knowledge gaps the present study intends to uncover. Finally, one may observe that SAL theory is contrasted with IP theory in some instances. I do not intend to pick on the theory. However, I think IP theory forms a default basis for comparison with SAL theory because the latter evolves from a dissatisfaction with the former by Marton, Säljö, and other colleagues.

2.3 Self-efficacy theory

2.3.1 Origin, ontology, and epistemology of the self-efficacy theory

Self-efficacy theory is deeply rooted in Bandura's agentic social cognitive theory which sees an individual's behavioural changes as consistently being regulated and modified by interacting with social factors in the environment whose feedback influences the next actions and outcomes (Bandura, 2001). The agentic social cognitive theory made an ontological paradigm shift from the traditional social cognitivism by rejecting the dualistic view of personal agency and social structure (Bandura, 2012). Thus, to social cognitive theorists, "the self is not split into object and agent; rather, in self-reflection and self-influence, individuals are simultaneously agent and object" (Bandura, 1997, p. 5). This assertion means when a person acts intentionally on things around in the environment, he or she becomes an agent. Almost concurrently, when he or she acts on self or engages in self-reflection, he or she becomes an object. Therefore, Bandura argued that both the personal agency and the social structure "function interdependently rather than as disembodied entities" (Bandura, 2012, p. 15).

An epistemological position that is fundamental to the agentic social cognitive theory is the concept of reciprocal determinism (Bandura, 2012). It is a perspective with which human functioning is viewed as a causal dynamic system of interaction between personal factors, behavioural factors, and environmental factors. Thus, perceived self-efficacy, which is a component of the personal factors in the dynamic system, is considered not to be a fixed construct. Instead, it changes accordingly with respect to changes in the system (Bandura, 2012). In specific terms, Borgonovi and Pokropek (2019) applied reciprocal determinism to research on mathematics perceived self-efficacy. They wrote, "reciprocal determinism describes the sets of relationships underlying the interactions between: (a) individuals' exposure to mathematics tasks, (b) mathematics self-efficacy beliefs, and (c) mathematics ability" (p. 269). Therefore, it can be argued that perceived self-efficacy is a task-specific construct that influences the performance of engineering students in a calculus task.

2.3.2 Perceived self-efficacy is a combination of confidence and estimations of expected outcomes

Apart from the task-specificity of perceived self-efficacy, another basic tenet of self-efficacy theory is that all psychological and behavioural changes occur as a result of modifications in the sense of efficacy or personal mastery of an individual (Bandura, 1977, 1982). In the words of Bandura (1977), “people process, weigh, and integrate diverse sources of information concerning their capability, and they regulate their choice behavior and effort expenditure accordingly” (p. 212). Thus, Bandura’s self-efficacy theory posits that explanations and predictions of psychological changes can be achieved through appraisal of perceived self-efficacy expectations of an individual. In other words, perceived self-efficacy is a combination of both outcome expectancy – credence that a given behaviour will or will not result to a given outcome – and self-efficacy expectancy – “the belief that the person is or is not capable of performing the requisite”, (Maddux, Sherer, & Rogers, 1982, p. 208). As such, one cannot wholly discern perceived self-efficacy from the expectancy-value theory (Wigfield & Eccles, 2002). However, self-efficacy theory places more emphasis on competence beliefs than the expectancy-value theory does (Leaper, 2011).

2.3.3 Perceived self-efficacy regulates the adoption of approaches to learning

Perceived self-efficacy has been documented to form a strong positive relationship between ‘challenging set goals’ and ‘commitment for its attainment’. In Bandura’s words “the stronger the perceived self-efficacy, the higher the goal challenges people set for themselves and the firmer is their commitment to them” (Bandura, 1993, p. 118). Perceived self-efficacy beliefs serve several purposes in regulating cognitive, motivational, affective, and decisional processes of an individual’s human functioning (Bandura, 2001, 2002). It is crucial to a learner as it stimulates the individual not to relent in completing difficult tasks despite hindrance. It makes the individual’s involvement very active and boosts morale to see to the attainment of a desirable outcome (Bandura, 1997, 2012). In a study involving undergraduate students following a biomechanics course in the United States, Wallace and Kernozek (2017) demonstrated how self-efficacy theory could be used by instructors to improve students learning experience and lower their anxiety towards the course. More so, Sheu et al. (2018) reported a meta-analysis study on the contributions of self-efficacy theory in learning science, mathematics, engineering and technology. More recently, Czocher, Melhuish, and Kandasamy

(2019) showed how interventions such as a mathematical modelling competition could be used to improve students' mathematics perceived self-efficacy.

Moreover, one may argue that perceived self-efficacy is not only a predictor of engineering students' learning outcomes in mathematics, but it also regulates the adoption of approaches to learning from two perspectives. First, since perceived self-efficacy regulates individuals' decisional processes and approaches to learning are parts of these processes (Biggs, 1993), then perceived self-efficacy could influence adoption of students' approaches to learning. Second, deep approaches to learning are motivated intrinsically, surface approaches to learning are motivated extrinsically, and perceived self-efficacy regulates motivational processes. Then, perceived self-efficacy should influence the adoption of approaches to learning though motivation. As such, I formulate the following hypotheses:

> *Hypothesis 5*: There is a causal effect of perceived self-efficacy on engineering students' ongoing approaches to learning a first-year calculus course.
>
> *Hypothesis 6*: There is an effect of perceived self-efficacy on engineering students' performance in a first-year calculus course.

2.3.4 Prior mathematics knowledge and perceived self-efficacy

The development of people's beliefs to complete a task in order to achieve a desirable outcome has been reported to have four primary sources of influence. These sources are enumerated as follows: "enactive mastery experience" – personal previous task-based achievement, "vicarious experience" – experience gained by monitoring peers or people around, "verbal persuasion" – complementary or contradictory feedback received from others, and "physiological and affective states" – physical or emotional situations during the behavioural changes (Bandura, 1997, p. 79; 2008). Personal experience on previous tasks exerts the most substantial influence among the four sources of perceived self-efficacy with successes consolidating a robust perceived self-efficacy and failure, on the other hand, weakening it (Bandura, 2008; Yurt, 2014). Within the context of mathematics learning, prior knowledge, among other sources of perceived self-efficacy, has been shown to have the highest impact on students' perceived self-efficacy in solving mathematics problems (e.g., Joët, Usher, & Bressoux, 2011; Zientek, Fong, & Phelps, 2019).

At this juncture, one may argue that Bandura's self-efficacy theory can also be embedded in the 3P model as it emphasises the effect of previous experience (a presage factor) on perceived self-efficacy (a process factor), which in turn influences students' performance (a product factor). As a result, I claim that prior mathematics knowledge influences perceived self-efficacy, which in turn affects engineering students' performance in mathematics. Therefore, I formulated the following hypotheses:

> *Hypothesis 7*: There is an effect of prior mathematics knowledge on perceived self-efficacy among first-year engineering students.
> *Hypothesis 8*: Perceived self-efficacy mediates the effect of engineering students' prior mathematics knowledge on their performance in a first-year calculus course.

2.4 Reflective critique

2.4.1 Implications of SAL theory to the present study

The SAL theory of Marton and Säljö using *phenomenography* coupled with some modifications and advancement by Biggs and others has provided theoretical structures for conceptualising students' approaches to learning in the present study. This statement is evident in the way the engineering student's approaches to learning have been defined to include motives, predispositions, styles, strategies used in adopting a process rather than mere cognitive activity as in the IP framework. Moreover, the classification of students' approaches to learning into 'surface' and 'deep' approaches has influenced the present study towards selection and validation of a measure for the constructs (Zakariya, Bjørkestøl, Nilsen, Goodchild, & Lorås, 2020). According to SAL tradition, approaches to learning are not fixed, they are motivated by intention and purpose, and they influence learning outcomes. These basic ideas of the theory have strengthened the present study in formulating some research questions and hypotheses. However, SAL theory is limited in scope as it only concerns approaches to learning and cannot be used to justify relations between perceived self-efficacy and learning outcomes in mathematics. This limitation of SAL theory led to the adoption of Bandura's self-efficacy theory.

2.4.2 Implications of self-efficacy theory to the present study

The conceptual understanding of perceived self-efficacy in the present study has been provided with a theoretical structure from Bandura's self-efficacy theory. Self-efficacy theory posits that all psychological and behavioural changes occur as a result of modifications in the sense of self-efficacy – the conviction to perform a task geared towards a desirable outcome, or personal mastery of an individual. Perceived self-efficacy is a task-specific construct. It has four sources of influence in which prior knowledge exerts the most substantial influence on the construct, and it affects learning outcomes in mathematics. These basic arguments of self-efficacy theory shaped the formulation of some research questions and hypotheses of the present study. Further, this theory has a strong instrumental role in developing the calculus self-efficacy inventory used in the present study, reported elsewhere (Zakariya, Goodchild, Bjørkestøl, & Nilsen, 2019).

2.4.3 Potential for networking SAL and self-efficacy theories

It is important to remark that the two theories used in the present are complementary to each other with a common aim of explaining differences in learning outcomes among higher education students. Complementarity in a sense that the self-efficacy theory addresses the aspects of perceived self-efficacy that the SAL theory cannot justify. Their ontological and epistemological positions on the constructs they address as being context-specific and influenced by social factors in a non-dualistic manner are potentials for networking these theories. There appears to be a common ground for the two theories in their paradigmatic research questions of either explaining or predicting the learning outcomes from students' factors, even though their methodological approaches are different in terms of phenomenography in SAL theory and quantitative methodology in self-efficacy theory. It can, therefore, be argued that these theories are not too distant from each to make a coherence argument for the present research.

Another shared characteristic of the two theories is that they consolidate the 'presage', 'process' and 'product' features of the 3P model. The 3P framework is necessary to anchor these two theories in a way that the eight hypotheses of the present study could form a uniform hypothesised conceptual model, as presented in Figure 2. Such that prior mathematics knowledge is embedded in the 'presage', ongoing learning approaches and perceived self-efficacy are embedded in the 'process', and performance of students on mathematics tasks is embedded in the ‘product’ components of the 3P model. However, there are other aspects of engineering students’ learning that are not captured by the two theories. Some of these aspects are collective dimensions and the students’ activity systems, teacher-student interaction, students’ mathematical discourse, contradictions, and tensions in learning first-year calculus course. In order to attend to these shortcomings, future (qualitative) research may be conducted using the activity theory, the anthropological theory of didactics, and commognition theory as theoretical approaches.

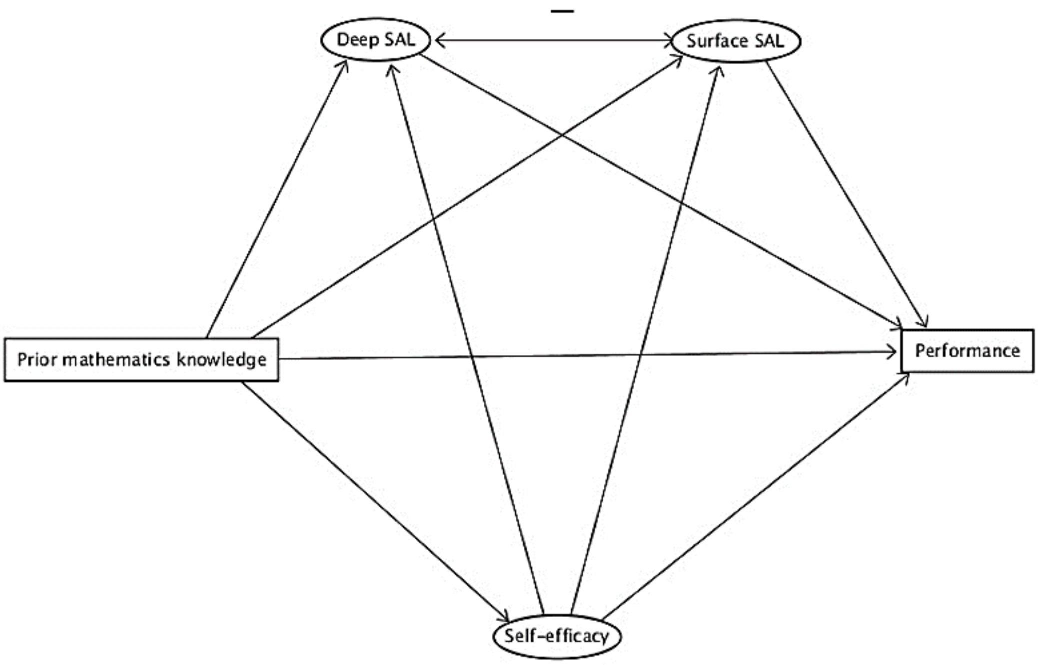


Figure 2. The Hypothesised conceptual relationships between research variables

Figure 2 shows the hypothesised relationships between the research constructs. The oval shapes represent unobserved (latent) variables, while the rectangles represent observed variables. Single-headed arrows indicate the directions of the hypothesised effects, and the double-headed arrow indicates a correlation. The figure shows that prior mathematics knowledge, deep approaches, surface approaches, and calculus self-efficacy are hypothesised to have direct effects on students’ performance in the first-year calculus course. Prior mathematics knowledge is hypothesised to have an indirect effect on performance in the first-year calculus course via calculus self-efficacy and approaches to learning. Further, calculus self-efficacy is hypothesised to have an indirect effect on student performance in the first-year calculus course via approaches to learning, and there is a negative correlation between deep and surface approaches. The correlation between deep and surface is negative because each student who adopts a deep approach to learning the calculus course is expected to have a low score on surface approach subscale and vice-versa.

2.5 Summary of the chapter

To conclude, the current chapter presents a brief introduction to frameworks in mathematics education research. This introduction was followed by my argument for justifying the appropriateness of adopting the conceptual framework in the present study rather than the practical or the theoretical frameworks. Primary tenets of the two theories that form the conceptual framework of the present study were discussed coupled with conceptualisations of approaches to learning and perceived self-efficacy. I demonstrate how the postulates from the two theories coupled with some insights from previous studies led to the formulation of my research hypotheses. Further, my reflective thoughts on the strengths and limitations of the two adopted theories with an argument for potentials on networking them were presented.

3 Methodology and methods

3.1 Research paradigm

There seems to be no agreement among philosophers, educators, and scientists on a precise definition of paradigm. In the classic book on the philosophy of science "*The structure of scientific revolutions"* by Thomas S. Kuhn, Margaret Masterman pointed out twenty-one different usages of paradigms in which Kuhn admitted that there are indeed twenty-two distinct usages (Kuhn, 2012). These distinct usages of paradigms have probably stemmed from its various conceptualisations across different fields of study. To bypass these controversies, one can refer to Kuhn's (2012) definition of paradigm at some points in his book as "the entire constellation of beliefs, values, techniques, and so on shared by the members of a given community" (p. 174). Therefore, a paradigm may be described as shared beliefs, values, and methods by a community of educators in relation to conducting scientific inquiries in the field. In the forthcoming subsections, rather than announcing a research paradigm (which is a difficult task for me) for the present study, I present my stances from the following perspectives:

- nature of reality of chosen attributes under study – ontological claims,
- relations between the researcher(s) and the attributes under investigation – epistemological claims, and
- procedures adopted by the researcher(s) to find out what can be known about the chosen attributes – methodological claims (Lincoln & Guba, 2005).

3.1.1 Ontological claims of the research paradigm

The ontology of the present study follows the lines of thought of critical realism with a notion that the chosen attributes - students' learning approaches, self-efficacy, prior mathematics knowledge and performance have separate existence that is independent of whether the researcher is thinking about them or not. It is acknowledged that there is an objective reality to each of these attributes. However, my apprehension of this reality is theory-laden, which is contrary to the empiricist view (and its social science version – positivism) that claims total apprehension of reality. This view also contradicts relativism that denies separate and absolute existence of such attributes and claims that multiple realities exist for each attribute. More so, I hold that individuals experience these attributes in different ways, either directly or indirectly observable. Moreover, these attributes can be operationalised and measured[4]. Direct apprehension of reality is a big claim which seems practically impossible when dealing with human subjects whose next actions are difficult to predict due to consciousness. However, it seems August Comte failed to acknowledge this fact when he advocated the adaptation of empiricism to studies in the field of social sciences. I also struggle with this idea, especially in the proper place to position mathematics education research. A clear-cut demarcation seems obscure as to whether mathematics education belongs to natural sciences or social sciences. This confusion has metamorphosised, in my opinion, into lack of an acceptable universal epistemology of research in the field.

3.1.2 Epistemological claims of the research paradigm

On the question of how to approach the research constructs? I approached, studied, and treated the research constructs as social phenomena that confront the researcher as external factors each with a distinct existence in an objective way and not constructed. Objectivity in this sense connotes an approach that reflects the true nature of the constructs to a large extent, free from biases and intersubjectivity, i.e., repeated observations by different researchers shall produce similar (not exact) data. It is acknowledged in the present study that complete objectivity is not feasible but can be approached as close as possible. I am neither in support of subjectivism as understood by the constructivists nor positivists who claim researchers should be neutral and devoid of all biases. That is, I hold a contrary perspective to the accurate apprehension of reality as claimed by the positivists and localised representation of realities as claimed by the constructivists. Thus, findings that emanate from the present study represent, supposedly, a close estimation of the relationships between the research constructs.

3.1.2 Logic of methodology of the research paradigm

The logic of methodology in the present study follows a scientific approach of hypothetical deductions as postulated by Popper (2002). The effects between and within the research constructs were hypothesised based on theories as explained in the conceptual framework section and literature before data collection. The data

---

[4]The issue of measuring the research constructs will be elaborated in the next section.

were collected through quantitative methods involving the use of survey instruments and tests. The hypotheses were then tested using critical descriptive and inferential statistics such as exploratory factor analysis, confirmatory factor analysis, item response theory, and structural equation modelling. These enable the researcher to achieve his inquiry aims of establishing the relationships between and within the research constructs. This methodology is considered appropriate because of its alignment with the ontology and epistemology of the present study, unlike the qualitative research methodology that is popular with relativism and subjectivism. Moreover, the use of structural equation modelling to evaluate the hypotheses requires large amount of data to ensure adequate measures of validity, reliability, significance, and effect sizes. Data from the large cohort of students could be too complex to analyse with the use of qualitative research methodology.

3.2 Conceptualisations of measurement

My statement in Section 3.1.2 that the research constructs can be operationalised and measured has provoked some reflections about what is measurement? Put differently, what is my perception of the word measurement? To address this question, I must allude to the fact that a thorough explanation of the word measurement is indefensible without a resort to philosophical theories on measurement. As such, I will highlight the conceptions of the measurement as articulated by three significant theories of measurement – classical theory (not classical test theory), representational theory, and latent variable theory – and justify my choice of upholding the latent variable theory of measurement.

3.2.1 The classical theory of measurement

The classical theorists (e.g., Michell, 1986) maintain a realistic ontological perspective of an attribute to measure while conceptualising measurement as “the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute” (Michell, 1997, p. 358, italics removed). A classic example is the length of a side of the tabletop. To estimate this length (or discover the ratio of the side of a tabletop’ length to a unit, such as a meter), one enumerates the number of units that make up a length which equals to that of the side of the tabletop. This procedure can be done quickly by holding a meter rule next to the side of the tabletop and reading off the equivalent units of length from it. As such, measurement, according to the classical theory, entails logical arguments that establish a *quantitative structure* for the attribute to be measured coupled with formulations of methods for a systematic numerical estimation of magnitudes. Therefore, for any attribute (e.g., length of the side of a tabletop) to be measurable, it must possess a quantitative structure, and such attribute is called a quantity (Michell, 1997).

An attribute is said to possess a quantitative structure if, most importantly, it satisfies the first axiom[5] by Hölder (1901) with an English translation by Michell and Ernst (1996). Moreover, the attribute, by extension, is commutative, associative and continuous (Michell, 1997). Hölder's first axiom demands that a quantitative attributive must be *divisible into divisibles*. For instance, the length of a pencil (or mass of an object) is divisible when the pencil is broken into smaller parts with each smaller part still regarded as a length of the pencil. In this case, the length, so also the mass of an object, satisfies Hölder's first axiom. More so, two or three broken lengths of the pencil (or bits of the mass of an object) can be put together to form a longer length, and both the order at which this combination takes place and its grouping are of no consequence to the resultant length. Thus, the length, so also the mass of an object, is both commutative and associative. Furthermore, a quantitative attribute must be continuous, i.e., *infinitely dense*. This means, for every two unequal broken lengths of a pencil, for instance, there must exist a third piece that is either greater than the first and less than the second or less than the first and greater than the second. It is important to emphasise that the conditions of quantitative structure apply to the essence of the attributes and the numerical estimations to quantify them.

The restriction of measurable attributes to quantities in the classical theory of measurement poses a challenge to any claim of measuring psychological attributes such as perceived self-efficacy. The simple question is, do psychological attributes possess quantitative structure? A simple answer is no because most psychological attributes such as perceived self-efficacy cannot be combined the same way quantities can be combined. We know that two objects can be added by aligning side by side or put together in a container to get a heavier object. However, it is

---

[5] Hölder's first axiom states that "[g]iven some quantity Q with levels $(a, b, c, \ldots)$, either (i) $a = b$, (ii) there exists c in Q such that $a = b + c$, or (iii) there exists c in Q such that $b = a + c$" (Markus & Borsboom, 2013, p. 22).

unknown how such additive property can be theoretically confirmed for perceived self-efficacy. Further, an object of mass 20 grams is two times heavier than an object of mass 10 grams. However, there is no evidence to support the claim that a person with a score of 20 on a test of prior mathematics knowledge is twice competent on the mathematics content knowledge that the test is designed to expose than a student with a score of 10 on the same test. Thus, at both the attribute and the score levels most psychological attributes, if not all, do not possess quantitative structure. Therefore, it is a fruitless exercise and self-contradictory to claim measurement of psychological attributes and yet be a loyalist of the classical theory of measurement (Michell, 1997, 1999).

3.2.2 The representational theory of measurement

The legitimacy of measuring psychological attributes was defended by some theorists (e.g., Stevens, 1946) with an introduction of the representational theory of measurement. Albeit this defence comes with a price. This price is a more liberal approach of extending the concept of measurement to include attributes that do not possess quantitative structures. This attempt gave rise to a broader conceptualisation of measurement as “the assignment of numerals to objects or events according to rule” (Stevens, 1946, p. 677). Stevens’ definition of measurement opened doors for different levels of measurement (nominal, ordinal, interval, and ratio) because there are different rules one can utilise while assigning numerals to objects. With Stevens’ conception of measurement, attention was shifted from the realist view of the attribute to be measured (as emphasised in the classical theory) to operationalist view that emphasises standardised procedures of assigning numerals to objects. The fact that Stevens’ representational theory emphasises rules in his conception of measurement makes his theory susceptible to criticisms. First, for creating an impression that anything is measurable by mere assignment of numbers, and second for neglecting the quantitative structure of the attributes to be measured (e.g., Michell, 1997, 1999; Trendler, 2009). Perhaps, these criticisms led to a more refined conceptualisation of measurement and the introduction of the axiomatic approach to representational theory.

The axiomatic approach to the representational theory of measurement retains the conceptualisation of measurement put forward by the Stevens (1946) but redefines the rules with the imposition of *isomorphism* on the representation (Luce & Suppes, 2002; Tversky, Krantz, Suppes, & Luce, 1971). Where, “an isomorphic representation is an assignment of numerals to objects such that every relation between the numerals has a parallel relation between the objects" (Markus & Borsboom, 2013, p. 35). As an example, suppose objects X and Y are assigned numbers x and y, respectively, and x is greater than y, then X is longer than or heavier than Y, and vice versa. This idea appears to be trivially understood. However, when it is applied to the measurement of psychological attributes such as perceived self-efficacy the story will drastically change. This is because psychological attributes are susceptible to measurement disturbances which make it challenging to achieve isomorphic representation as propagated by the representational theory (Cliff, 1992; Trendler, 2009). I will close this section with an opinion of Trendler (2009) about the proposal of axiomatic representational theorists on measuring psychological attributes. He argued that "*they* [psychological attributes] *are neither manipulable nor are they controllable to the extent necessary for an empirically meaningful application of measurement theory. Hence, they are not measurable*" (p. 592, italics in the original).

3.2.3 The latent variable theory of measurement

The clash of paradigms between the classical and the representational theories of measurement seems to be resolved by the latent variable theory of measurement (Borsboom, 2008). Measurement, according to the latent variable theory "involves determining the position of people in a latent space on the basis of sets of fallible indicators" (Markus & Borsboom, 2013, p. 68). As such, all psychological attributes are assumed to be unobserved variables (latent variables) whose variability constitutes a common cause for the covariation in respondents' scores on observed variables (indicators). Another crucial assumption in the latent variable theory is *local independence*. This assumption holds that if the latent variable does not vary then the covariation in respondents' scores on observed variables will vanish. That is, "an item measures a particular attribute only if differences on the attribute cause differences in the item scores" (Markus & Borsboom, 2013, p. 84). The latent variables, e.g., perceived self-efficacy, are mostly continuous attributes that are not directly observable. At the same time, indicators are scale items designed to expose the latent variable(s) based on substantive theory (Kline, 2016). The substantive theory, e.g., self-efficacy theory, provides the structure of the attribute (as discussed in conceptual framework section) and there is no need to assume quantitative structure for such psychological attributes (Markus & Borsboom, 2013).

On the one hand, the latent variable theory upholds the realist view of psychological attributes (as in the classical theory) but rejects the quantitative structure (as in the representational theory). On the other hand, the latent variable theory does not subscribe to the isomorphic representation and acknowledge that the functional relationship between the latent variables and indicators are susceptible to disturbances which can be controlled during modelling (Borsboom, Mellenbergh, & van Heerden, 2003; Markus & Borsboom, 2013). The latent variable theory is adopted in the present study for several reasons. Some of the reasons include its recourse to substantive theory to understand the structure of each psychological attribute, its rich testable causal assumptions, and its robustness to noisy data (Antonakis et al., 2010; Bollen & Pearl, 2013). Examples of models that are based on the latent variable theory are exploratory factor analysis (not to be confused with principal component analysis), confirmatory factor analysis, item response theory, and structural equation modelling. These models form the essential tools of analysis in the present study. More about these models will be presented in the forthcoming sections, including their implications to test validity and reliability. In the meantime, I will present some specifics of research methods before returning to latent variable models.

3.3 Research design

The present study adopts a cross-sectional survey design through which data are collected using, for the most part, online questionnaires, and a test from a large cohort of university students. Survey design is considered appropriate for this study as it grants the opportunity to quantify the factors that influence students' performance in mathematics. Further, it facilitates the investigation of inherent causal relationships within the research constructs using some substantive theories and advanced statistical analyses.

3.4 Sample of the study

The sample for the present study was collected in two phases. The first phase sample (pilot study) comprised first-year students who consented to take part in the project and had followed an introductory mathematics course in autumn 2018. The second phase sample (main study) comprised first-year engineering students who consented to take part in the project and followed an introductory mathematics in autumn 2019. I focus on the first-year engineering students in the present study for several reasons. First, it will avail me an opportunity to avert incoherence findings as it happened to Maciejewski and Merchant (2016) when they included students from other years of study in their study. Second, it is assumed that prior mathematics knowledge of the students may be assessed adequately in the first year of study. Third, consistent with the task-specificity of both perceived self-efficacy and approaches to learning mathematics, as highlighted in the conceptual framework, I focus on students following a common course. As such, I delimit my study to engineering students because they form the largest population of students enrolled on a common mathematics course at the university.

3.5 Operationalisations and measures

3.5.1 Operationalisation of approaches to learning

To operationalise approaches to learning to mathematics, it is necessary to highlight two crucial elements in the conceptualisations of these constructs. These crucial elements are *motives* and *strategies* (Biggs et al., 2001). For deep approaches to learning mathematics, the motive is to develop a conceptual understanding of the learning content while successful performance on the course becomes a by-product. As such, students devise lots of strategies to actualise this intention while learning mathematics. In contrast, the motive behind the adoption of surface approaches to learning mathematics is to pass the course with as little work as possible. In which case, students devise several strategies to actualise this intention while learning mathematics. Table 1 presents elaborated specifics of the motives and strategies behind the adoption of both deep and surface approaches to learning mathematics.

Table 1. Specifics of the operationalisation of approaches to learning mathematics

Deep approaches to learning	Surface approaches to learning
Excitement about new mathematics topics and devotion of spare time to develop a proper understanding of the topics.	The aim is to pass the course with limited work done.
Self-testing on crucial mathematics topics to develop mastery of the subject matter	Gross use of memorisation techniques with less care for developing a proper understanding of the mathematics content
Study hard for mathematics because of personal interest and feeling of satisfaction in the subject	Remembering answers to plausible examination questions is considered the best method to pass the examinations.
High preparation for mathematics classes with unanswered questions during students' self-study	Thinking that in-depth preparation for classes or study of mathematics topics is unnecessary, it wastes time and confusing.
Exploration of suggested readings for the course to develop more calculation skills	Self-confinement to class materials with a thought that is unnecessary to solve extra mathematical tasks

3.5.2 The measure of approaches to learning mathematics

Consistent with the conceptual framework of the present study, the revised two-factor study process questionnaire (R-SPQ-2F) was identified as the best among several measures that are developed to assess students' approaches to learning. Apart from being developed based on SAL theory, its high psychometric properties, short length, and ease of score interpretations give it more advantages over similar measures, e.g., study skills inventory for students (ASSIST) and revised approaches to studying inventory (RASI). The R-SPQ-2F conceptualised and operationalised approaches to learning into deep and surface approaches and measured each approach with ten items on a five-point Likert scale. Sample items on the R-SPQ-2F are presented in Table 2. R-SPQ-2F is a revision of earlier versions that date back to 1987 by Biggs et al. (2001). It has received global acceptance among researchers, translated and validated in several languages. However, these cross-cultural adaptations have equally provoked heated debates about cultural sensitivity of R-SPQ-2F (e.g., Immekus & Imbrie, 2010; López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014). For this reason, R-SPQ-2F was translated to Norwegian, localised to mathematics (appendices A and B) in line with the context-specificity of the construct, and validated for its construct validity, discriminant validity and reliability based on some procedures that will be described in the next section. The results of these validation studies are well documented and will be presented in the next chapter (Zakariya, 2019; Zakariya, Bjørkestøl, et al., 2020). Following the latent variable theory, the deep and surface approaches to learning are latent variables that manifest through or are exposed by their respective items on the R-SPQ-2F.

Table 2. Sample items of the R-SPQ-2F

	Motive	Strategy
Deep approaches to learning	"I find that at times studying gives me a feeling of deep personal satisfaction."	"I find most new topics interesting and often spend extra time trying to obtain more information about them."
	"I work hard at my studies because I find the material interesting."	"I test myself on important topics until I understand them completely."
Surface approaches to learning	"My aim is to pass the course while doing as little work as possible."	"I only study seriously what is given out in class or in the course outlines."
	"I find I can get by in most assessments by memorising key sections rather than trying to understand them."	"I find the best way to pass examinations is to try to remember answers to likely questions."

Note. All items are reprinted from "The revised two-factor Study Process Questionnaire: R-SPQ-2F" by J. Biggs, D. Kember and D.Y.P. Leung, 2001, *British Journal of Educational Psychology*, *71*(1), p. 148 (https://doi.org/10.1348/000709901158433). Copyright 2001 by the British Psychological Society. The 19-item Norwegian version of R-SPQ-2F and its English back translations are available in Appendix 1 and Appendix 2, respectively.

3.5.3 Operationalisation of perceived self-efficacy

A good number of educators have argued that the best measures of perceived self-efficacy with high predictive power of students' performance in mathematics are task-specific measures (e.g., Kranzler & Pajares, 1997; Pajares & Miller, 1995). As such, to operationalise perceived self-efficacy, it is as well necessary to highlight two crucial elements in the conceptualisation of the construct as informed by the self-efficacy theory. These crucial elements are *confidence* and an *estimation of the expected outcome* (Bandura, 2006). That is, for a student to be described of possessing calculus perceived self-efficacy, for instance, he or she must express his or her convictions of solving some presented calculus tasks and must provide an estimate of an expected score or a percentage on a metric scale defined by the researcher.

3.5.4 The measure of mathematics perceived self-efficacy

Surprisingly, an extensive search of the literature revealed a lack of perceived self-efficacy measure specifically designed to expose student's perceived self-efficacy on year-one calculus tasks. Therefore, with the operationalisation of perceived self-efficacy as highlighted above coupled with Bandura's (2006) guidelines, I developed a measure of calculus perceived self-efficacy, namely, calculus self-efficacy inventory, using the following steps:

1. I extracted 15 exam-like questions from the old examination papers of the target introductory calculus course. The extracted questions are distributed across topics from the curriculum course content (e.g., limits, functions, differentiation, and integration).

2. I prepared these questions into an inventory in which students are to rate their confidence in solving each task and by how much on a scale of 100 points.

3. Then, the inventory was pilot-tested and validated for its psychometric properties. The results of this validation study are well-documented and will be presented in the next chapter (Zakariya et al., 2019).

The calculus perceived self-efficacy is a latent variable that manifests through or is exposed by the items of the self-efficacy inventory. Figure 3 shows some sample items and the stem question of the calculus self-efficacy inventory. The full inventory is available in Appendix 3 and Appendix 4 for both the Norwegian and the English versions, respectively.

Below you will find a set of different tasks from the Calculus syllabus. In the column "Confidence" you are asked to state how much confidence you have that you could manage to solve each task – **here and now**. Please provide your level of confidence with a number from 0 to 100, using the scale below:

0	10	20	30	40	50	60	70	80	90	100
No possibility that I could solve this				Moderately confident that I could solve this					Totally confident that I could solve this	

Task	How confident are you that you can solve each of these problems **right now**?	Confidence (0 – 100)
1	Calculate $(1-i)^{666}$	
2	Calculate the value of the limit $$\lim_{x\to 1}\frac{1-\cos(1-x^2)}{x^2-2x+1}$$	
3	Given the function $f(x)=\ln(2x^2-3x+2)$. What is the domain of the function f?	
4	Given the function $f(x)=x+e^x$. What is the value of c that satisfies the result of the **Mean Value Theorem** in the interval $[0,1]$?	

Figure 3. Sample questions of the final version of the calculus self-efficacy inventory

3.5.5 Operationalisation and measure of prior mathematics knowledge

The prior mathematics knowledge of first-year engineering students varies distinctively depending on the types of mathematics – practical mathematics (P1 and/or P2), and theoretical mathematics (S1, S2, R1 or R2) – they followed in their respective upper secondary schools. To find a common baseline in assessing this diverse prior mathematics knowledge without compromising the entry requirements to engineering programs, I adopted and validated the Norwegian national mathematics test (NMT) and used it in the present study. The NMT is a test that is owned and conducted biennially for the past two decades by the Norwegian Mathematical Council. The Mathematical Council designed the test to assess pre-university knowledge of mathematics of undergraduate students on entry to degree programs across universities in Norway. Therein, 16 items are drawn from the lower secondary school mathematics curriculum in which the students are to solve within 40 minutes. Some of the items are standard multiple-choice questions while others are short answer open-ended questions. Prior mathematics knowledge in all the analyses of the present study was treated as a latent variable that is manifested through or is exposed by the items of the NMT. The item quality (in terms of difficulty and discrimination indices), validity, and reliability of the NMT are studied and well-documented (Zakariya, Nilsen, Goodchild, & Bjørkestøl, 2020a). I did not include sample items on the test in the present study because permission to do so has not been granted by the Norwegian Mathematical Council, and this because the test is reused, to show trends of students' competencies over several years.

3.5.6 Operationalisation and measure of performance in mathematics

Students' performance in mathematics was operationalised and measured in the present study with their respective cumulative final scores in an introductory year-one mathematics course. Consistent with the literature (e.g., Cano et al., 2018), I argue that these scores are the most suitable measure of students' performance in the course. My argument is based on a premise that the scores are reflective of the overall achievement on university regular assessment that is common to all the students.

3.6 Test validity and reliability

3.6.1 Background to factor analysis

The test validity of the measures used in the present study was investigated using factor analysis procedures. Factor analysis (FA) is a statistical tool that has gained wide acceptance among educationists over many decades in developing research measures, e.g., questionnaires, and tests (DiStefano & Hess, 2005). Exploratory and confirmatory factor analyses are two examples of FA with a shared origin called the common factor model. Because the common factor models are based on latent variable theory, constructs, e.g., perceived self-efficacy and approaches to learning mathematics are unobserved variables that are operationalised by some scale items/indicators such that covariances of scores on these items are reflective of the constructs. As such, the test validity is the degree at which the covariation in the observed item scores reflect their common cause (the unobserved/latent construct). Exploratory factor analysis (EFA) seeks to explain the variability by optimally expressing each indicator as a linear function of a unique factor and one or more common factors such that a minimal number of factors are identified. EFA is usually used at the early stage of scale development. On the other hand, confirmatory factor analysis (CFA) is used at a later stage of scale development to validate theorised relationships between indicators and factors of such measures. Thus, CFA constitutes a test of validity in the latent variable perspective. For instance, consider the measurement model of a Norwegian language validation of the R-SPQ-2F used in the present study, as shown in Figure 4.

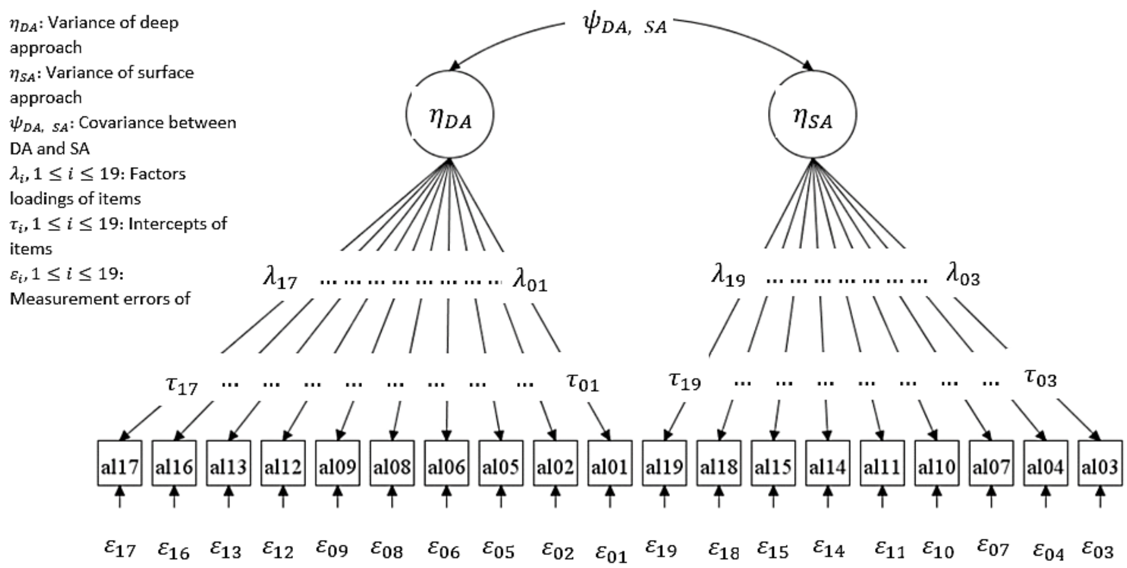


Figure 4. The measurement model of R-SPQ-2F (Norwegian version)
Note. The Norwegian version of the R-SPQ-2F contains 19 items as opposed to the 20 items in the original questionnaire.

The accompanied set of equations to the measurement model in Figure 4 are presented in Equation 1. Even though, these equations share some assumptions (e.g., linearity, i.e., a unit increase or decrease in η_j brings a unit change in al_i, constant coefficients, i.e. $\lambda_i's$ are assumed to be equal across every individual of the population, and so on) with regression equations (REs) they are substantially different from the REs in two ways. The first way concerns a causal assumption inherent in Equation 1. For each equation in (1), there is a causal effect of η_j on al_i whose magnitude is measured by λ_i such that the positions of η_j and al_i are not interchangeable in Equation 1 unlike in regression that assumes λ_i to be a mere correlation coefficient. The second way concerns the orthogonality assumption on $\varepsilon_i's$, i.e., $\varepsilon_i's$ have a mean of zero and neither correlate with each other nor with $\eta_j's$ in REs. However, orthogonality in the context of structural equations may or may not be assumed depending on model theorisation, literature, valid claims, and other sources that the researcher can rely upon for his or her arguments (Bollen & Pearl, 2013). Further, the $\varepsilon_i's$ are conceptually different from those of REs in that they incorporate errors that emanate from potential predictors of al_i which are not accounted for in the measurement model. It is important to remark that these assumptions form the building blocks of the models used in the present study, and a rejection of a model challenges the validity of one or more of these assumptions. On the other hand, an acceptance of a model strengthens the plausibility of these assumptions.

$$al_i = \tau_i + \lambda_i \eta_j + \varepsilon_i, 1 \leq i \leq 19, \qquad (1)$$

$$j = DA \; if \; i \; falls \; on \; deep \; approach \; else \; j = SA$$

Basically, CFA predicts a variance-covariance matrix $\sum$ from raw input data, sample correlation matrix or sample variance-covariance matrix (S) and minimises the residual matrix ($S - \sum$) to achieve a good fit. There are several estimators used for this procedure, e.g., maximum likelihood (ML), weighted least square mean and variance adjusted (WLSMV), and so on, depending on several assumptions of normality, levels of measurement, and presence of missing data. The degree of a good fit is assessed by chi-square (χ^2) statistics with a null hypothesis, $S = \sum$, coupled with some indices of the goodness of fits. Hence, a non-significant χ^2-value indicates a good fit of the model (Brown, 2015). However, researchers have contended that due to the large sample size required to conduct CFA, a small difference in the residual matrix could make χ^2-value to be significant which could erroneously lead to a rejection of a good model (Prudon, 2015). To ameliorate this problem, methodologists have introduced some goodness of fit (GOF) indices.

3.6.2 The goodness of fit indices

The GOF indices are used to assess the global fits of hypothesised measurement models against the collected data. Popularly reported GOF indices in education research are: TLI-Tucker-Lewis index (Tucker & Lewis, 1973), RMSEA-root mean square error of approximation (Steiger and Lind, 1980 in Steiger, 2016), SRMR-standardized root mean square residual (Jöreskog & Sörbom, 1988), and CFI-comparative fit index (Bentler, 1990). Both TLI and CFI are examples of comparative fit indices or incremental fit indices (Hu & Bentler, 1998). These indices examine the fit of the predicted or implied matrix $\sum$ by comparing its χ^2 -value with that of a nested baseline model, equations 2 and 3. The values of CFI ranging from 0.00 to 1.00 while TLI may assume values out of this range. Hence, TLI is usually termed non-normed incremental fit index. For both CFI and TLI, a value 1.00 indicates a perfect model fit while values close to or greater than 0.90 indicate a good fit (Bentler, 1990; Hu & Bentler, 1999).

$$CFI = 1 - \frac{\chi^2_{implied} - df_{implied}}{\chi^2_{baseline} - df_{baseline}} \qquad (2)$$

$$TLI = \frac{\left(\frac{\chi^2}{df}\right)_{baseline} - \left(\frac{\chi^2}{df}\right)_{implied}}{\left(\frac{\chi^2}{df}\right)_{baseline} - 1} \qquad (3)$$

RMSEA is a parsimony correction fit index which seeks to examine the model fit of a sample matrix (S) by including a 'penalty function' while favouring a model with "fewer freely estimated parameters" (Brown, 2015, p. 71). It is less dependent on the residuals, and it is computed using χ^2-value, degree of freedom and N-sample size, Equation 4. RMSEA index ranging from 0.00 and has no upper bound with a value 0.00 signifying a perfect model fit. A cutoff RMSEA value of less than or equal to 0.06 was proposed by Hu and Bentler (1999) for a good model fit. Other experts (e.g., Browne & Cudeck, 1992) have proposed RMSEA values between 0.00 to 0.05 and 0.05 to 0.08 as depicting good, and an adequate model fits respectively. A model with RMSEA value between 0.08 to 0.10 is characterised as having a "mediocre fit" while models with RMSEA values greater than 0.10 should be rejected (MacCallum, Browne, & Sugawara, 1996).

$$RSMEA = \sqrt{\frac{(\chi^2 - df)/N}{df}} \tag{4}$$

SRMR stands out among the fit indices as it is the only index that does not involve χ^2-value in its computation. Instead, SRMR directly estimates the discrepancy between input sample matrix and predicted matrix. It is calculated by taking square root of the ratio of the corresponding sum of squares of the residuals to the number of residual, p(p+1)/2 where p is the number of scale items, Equation 5. It evaluates the model fit at an absolute level. Hence, an example of absolute fit indices. SRMR values ranging from 0.00 to 1.00 with a value of 0, indicating a perfect fit. A value less than or equal to 0.08 was suggested by Hu and Bentler (1999) as an indicator of a good fit. In practice, methodologists and researchers do not take the cutoff values of GOF indices as a rule of thumb. In fact, a close look at the work of Hu and Bentler (1999) revealed that their cutoff criteria are not generalisable especially when other estimators apart from ML are used and more than five indicators per factors are involved (Marsh, Hau, & Wen, 2004). Further, Hu and Bentler (1999) criteria have been considered unrealistic for most social sciences research, especially when the data involve ordinal scales with multiple violations of assumptions (Marsh et al., 2004). It is therefore helpful, and of course, the criteria adopted in the present study, to utilise a combination of these indices with some relaxations in cutoff values. Further, I assessed the local fits of hypothesised models against the collected data by using significant statistics of indicator factor loadings, effect sizes, and interpretability of other parameter estimates such as effect weights and residuals.

$$SRMR = \sqrt{\frac{sum\ (S_{ij} - \Sigma_{ij})^2}{p(p+1)/2}} \tag{5}$$

3.6.3 Structural equation modelling and causation

A statistical tool that takes measurement model analyses (EFA and CFA) to the next level of providing empirical evidence for ‘causes and effects’ between research variables is the structural equation modelling (SEM). SEM approach to modelling has been applied widely in educational studies and proved compelling in given potential cues to causal relations between latent factors (e.g., Roick & Ringeisen, 2018). However, it has been equally criticised for its "erroneous" causal claims from both ontological and epistemological perspectives. The ontological criticisms of SEM stem from the conception of reality as it relates to human characteristics. That is, such reality is very complex, and SEM attempts to reduce it to a set of linear equations lack validity. Critics argued that no amount of statistical tools (not limited to SEM) could accurately describe reality (e.g., Rogosa, 1987). This criticism to me seems valid, especially for those who claim direct apprehension of reality is possible in social science research. However, it has previously been mentioned while describing the ontology of the present research that direct apprehension of reality is not claimed. Hence, this criticism is relatively superficial with regards to the present study. Also, multiple theory-driven indicators that are used to operationalise each construct of the present study, to a large extent, give reinforcement to the plausibility of their common factors.

Critiques from the epistemological perspective revolve around the validity of SEM as regards to causal claims from nonexperimental research data. Critics argued that the necessary basic conditions of causal claims, e.g., correlations between variables, controlling extraneous variables, and establishing antecedents seem unrealistic with the adoption of SEM (e.g., Freedman, 1991). One could approach this criticism from many angles. First, it is essential to say in clear terms what SEM aims to achieve. According to Bollen and Pearl (2013), "SEM is an inference engine that takes in two inputs, qualitative causal assumptions and empirical data, and derives two logical consequences of these inputs: quantitative causal conclusions and statistical measures of fit for the testable implications of the assumptions" (p. 309). Thus, SEM is not looking for or discovering causal relationships from mere correlations. Second, the issue of causality transcends the boundary of SEM and extends to designs of the study. For instance, a longitudinal SEM study will, to a large extent, account for antecedents between variables more than a cross-sectional SEM study. In a similar manner, a quasi-experimental SEM study will control extraneous variables better than a longitudinal SEM study. Third, the present study does not aim at discovering causal relations between the research variables. Instead, with the help of SEM's high precision, it is envisaged that compelling potential cues to causal effects between the variables will be achieved. This high precision of SEM and its robustness to assumption violations give it several advantages over other statistical analyses, e.g., multiple regression, in exposing causal effects between research variables. According to Bullock, Harlow, and Mulaik (1994), SEM offers more than other statistical analyses in its:

> ability to analyse direct and indirect effects, assess both measurement and prediction error, allow multiple measures to represent latent variables, and provide simultaneous estimation of measured and structural relations in a complex, integrated mathematical model. (p. 262)

Another set of threats to the validity of SEM causal estimates could stem from endogeneity problems, where some potential causes of the 'effects' are not included in the model. I briefly refer to Antonakis et al. (2010) who have identified and provided solutions to 14 sources of endogeneity problems in a nonexperimental SEM study. Of relevance to the present study are those threats to validity with sources from "omitted variable" (e.g., exclusion of "fixed effects"), "measurement errors", "common-method variance" (e.g., gathering of data on approaches to learning mathematics and calculus self-efficacy at the same time), "model misspecification", and "inconsistent inference"(Antonakis et al., 2010, p. 1091). I acknowledge these threats to validity in the present studies, and I minimise them by following recommended solutions such as the use of instrumental variables to account for errors due to omitted variables, the inclusion of measurement errors in the model, and use of robust estimators.

3.6.4 Reliability

Reliability concerns the consistency of an instrument at measuring what it is designed to measure. In technical terms, the reliability of a composite score (Y) obtained from a measure is the ratio of the true variance of Y to the total variance of Y. A widely used index of reliability for the past six decades is the Cronbach alpha coefficient. More recently, there have been heated debates among methodologists on the appropriateness of using Cronbach's alpha coefficient in estimating the reliability of ordinal scale data (e.g., Schmitt, 1996; Sijtsma, 2009). Some of these debates have been provoked by gross misuses and misinterpretations of Cronbach's alpha especially in the presence of excess kurtosis and skewness, violation of tau-equivalent assumption, presence of correlated errors, and non-continuous item level of measurement that are inherent in ordinal data (e.g., Sijtsma, 2009). To avoid these challenges, alternative indices of reliability have been proposed, e.g., ordinal coefficient alpha, Omega, Beta, and H coefficients, and GLB-greatest lower bound coefficients, for estimating the reliability of ordinal scales (e.g., Raykov & Marcoulides, 2016; Zinbarg, Revelle, Yovel, & Li, 2005).

The omega coefficient and its extensions to the multidimensional scale using a latent variable approach by Raykov and Marcoulides (2016) were mostly used in the present study. The reliability indices that are based on the latent variable approach have been shown empirically to perform better than the Cronbach's alpha estimates under violations of multiple assumptions which are inherent in the ordinal data of the present study (e.g., Gadermann, Guhn, & Zumbo, 2012; Raykov & Marcoulides, 2016; Zumbo, Gadermann, & Zeisser, 2007). Simplified formulae adapted for the present research, which involves a unidimensional scale with correlated errors and a two-factor multidimensional scale without correlated errors are presented.

$$r_{RM} = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 + \sum_{i=1}^{n} V_i + 2 * \sum_{i=2}^{n} \sum_{j=1}^{i} V_{ij}} \quad (6)$$

$$r_{RM} = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2 + \left(\sum_{j=1}^{m} \lambda_j\right)^2 + 2 * F_{12} * \left(\sum_{i=1}^{n} \lambda_i\right) * \left(\sum_{j=1}^{m} \lambda_j\right)}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 + \left(\sum_{j=1}^{m} \lambda_j\right)^2 + 2 * F_{12} * \left(\sum_{i=1}^{n} \lambda_i\right) * \left(\sum_{j=1}^{m} \lambda_j\right) + \sum_{i=1}^{n} V_i + \sum_{j=1}^{m} V_j} \quad (7)$$

In both equations 6 and 7, r_{RM} is the Raykov and Marcoulides' coefficient with values ranging from 0 to 1 that is indicative of item internal consistency from weakest (0) to strongest (1). λ_i's and λ_j's are (standardised) factor loadings of the subscale indicators, V_i's and V_j's are (standardised) unique variances, V_{ij}'s are error covariances between indicators i and j, and F_{12} is the (standardised) covariance between factors 1 and 2.

3.7 Data collection and analysis

The data used for the present study were collected in two phases. Phase one (pilot study) data were collected in spring 2019 with the aim of validating R-SPQ-2F and developing CSEI. Every year-one student who consented to take part in the project completed both the R-SPQ-2F and CSEI. It involved two data sets composed of

234 engineering and economics students as well as 253 engineering students that followed a first-year mathematics course. The 234-data set was used to develop the CSEI using EFA in FACTOR program. It was also used to establish a relationship between calculus self-efficacy and approaches to learning. Further, the 253-data set was used to validate the Norwegian adaptation of the R-SPQ-2F. The method of analysis of R-SPQ-2F followed a series of confirmatory factor analyses using Mplus 8.3 program.

Phase two data collection took place in autumn semester 2019 at two instances with the aim of examining the hypothesised relationships between the research constructs. Instance one of phase two data collection took place at the beginning of the semester in which all year-one engineering students who consented to take part in the project completed R-SPQ-2F (these data were not used in the final analyses) and took the NMT. Instance two of phase two data collection took place after the mid-term break towards the end of the semester. Both R-SPQ-2F and CSEI were administered. I deliberately delayed the administration of the CSEI for the purpose of making it close to the end of the term. This delay is necessary to increase the predictive power of CSEI as a substantial part of the calculus syllabus would have been completed at the time. The reason being that the CSEI was developed based on exam-like questions from the calculus curriculum, as discussed in the previous section. The administration of both R-SPQ-2F and CSEI at the same time could constitute a common-method variance endogeneity problem and threatens the validity of the model estimates. However, this problem was acknowledged and was addressed during the final modelling process. The structural equation modelling was used to analyse the data such that the research questions, as well as the hypotheses as presented in Figure 2, are addressed accordingly.

3.8 Ethical issues

Prior to the data collection of the present study, I, together with my supervisor team, visited the students in their classrooms to inform them about the purpose of the study and seek their consent. The content of the students’ consent centred around making connections between their final grades in their mathematics courses and data collected through the surveys. It was made clear to the students that we will neither store their personal information nor keep any information that could allow anyone to identify them. We only need their permission to make the connection of data, because that can only be done if there are individual identifiers, e.g., names and student numbers. As soon as this is done, all means of identification will be removed. For additional security and confidentiality in the process, the connection of data was carried out by administrative staff in the division of student and academic affairs, and all means of identification were removed before the data sets were made available to the research team. All data used by the researcher were anonymised, and any publication from the research in journals, conference presentations or other means did not include anything from which participants may be identified. Only the staff in the division of student and academic affairs have access to the full data sets, and they are bound with strict rules of confidentiality.

Further, the students were informed that taking part in the project is voluntary. If anyone chooses to take part now and decide later to withdraw his/her consent, it can be done without giving any reason by writing to his/her teacher. There was no adverse consequence for anyone who does not wish to participate or later withdraws his/her consent. We ensured strict compliance with personal data protection as regulated by the Norwegian Centre for Research Data (NSD). We applied for and received NSD approval to proceed with data collection provided the students give their consent.

3.9 Summary of the chapter

This chapter attempts to articulate the ontological, epistemological, and related methodological issues of the present study. The concept of measurement was presented from three major measurement theories, coupled with arguments for adopting latent variable theory. I delved into the specifics of operationalisations and measures of the research constructs. Thereafter, I discussed in detail the test validity and reliability from the perspective of the latent variable theory. Then, I followed up the validity issues with the procedures for data collection and analyses. Finally, I presented some ethical considerations with regards to the principle of informed consent, human subject protection, personal data protection, and legality.

4 Overview of papers

The present chapter gives an overview of the eight papers that are put together to form the present dissertation. Herein, I emphasise the research aims, research methods, and the main results for each paper. Figure 5 provides an integrative view for navigating through the present chapter. In the figure, I present the links from the test validation studies to the structural validation studies. More so, I also present the links between each paper and the corresponding research question and hypotheses each paper is purported to address. Some peculiar abbreviations in the figure are ‘RQ’, for the research question, and ‘Hyp.’, for the hypothesis.

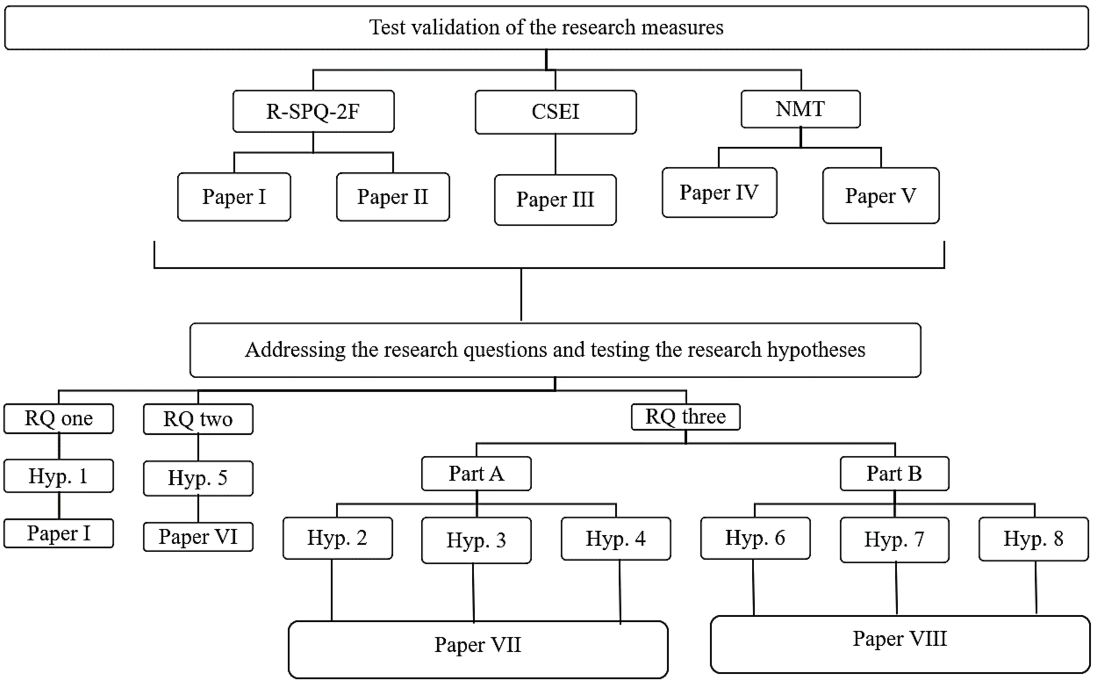


Figure 5. An integrative overview of papers in Chapter Four

4.1 Paper I and Paper II: Validation of approaches to learning questionnaire

Paper I: Zakariya, Y. F., Bjørkestøl, K., Nilsen, H. K., Goodchild, S., & Lorås, M. (2020). University students’ learning approaches: an adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Education Evaluation*, 100816. doi:10.1016/j.stueduc.2019.100816

Paper II: Zakariya, Y. F. (2019). Study approaches in higher education mathematics: Investigating the statistical behaviour of an instrument translated into Norwegian. *Education Sciences, 9*(3), 191. doi:10.3390/educsci9030191

4.1.1 Research aim

Studies reported in both Paper I and Paper II were aimed at addressing the first research question[6] and testing Hypothesis 1[7] by investigating the prevalence of deep and surface approaches to learning mathematics among first-year engineering students. Simultaneously, the studies were also aimed at evaluating the construct validity and reliability of the R-SPQ-2F before using the measure in the main study.

4.1.2 Research method

Consistent with the latent variable theory of measurement, I used a series of confirmatory factor analyses with a weighted least square mean and variance adjusted (WLSMV) estimator to establish the construct validity of R-SPQ-2F. Several proposed and hypothesised models by Biggs et al. (2001) and other researchers were evaluated and tested against my collected data. I assessed the plausibility of the models by using the global and local fit statistics, as highlighted in section 3.6.2 to ascertain evidence of construct validity of the measure in the Norwegian context. Similarly, I investigated the reliability of the measure using the latent variable approach, as presented in section 3.6.4.

4.1.3 Main results

The results of Paper I confirmed that two-latent factors (deep and surface approaches) are responsible for the covariation of the students' item scores on the measure. On the one hand, the results show that deep and surface approaches are prevalent among first-year engineering students. Thus, the results address the first research question of the present study and confirm the plausibility of Hypothesis 1. On the other hand, the results provide evidence for the construct validity of R-SPQ-2F. However, sufficient evidence of construct validity was achieved after

---

[6] Recall that the research question one is: Do approaches to learning mathematics differ with respect to the prevalence of deep and surface approaches among first-year engineering students?

[7] Recall that Hypothesis 1 is: There are differences in calculus learning approaches among first-year engineering students in terms of the prevalence of deep and surface approaches.

deleting one item[8] from the original measure. As such, the Norwegian version of the R-SPQ-2F contains ten items on the deep approach subscale and nine items on the surface approach subscale of the measure. In a search for more validity evidence for the constructs exposed by the R-SPQ-2F, I evaluated and compared ten different hypothesised R-SPQ-2F models with my established model in Paper I. The findings of these comparisons were reported in Paper II. Therein, the results show that the best explanatory model of the R-SPQ-2F was the 10-item deep and 9-item surface R-SPQ-2F that I established in Paper I. Furthermore, I got a reliability index of .81 for the deep approach subscale, an index of .72 for the surface approach subscale, and an index of .63 for the whole measure. The findings of these studies, on the one hand, reinforce my confidence to proceed with the use of the R-SPQ-2F in the main study. On the other hand, they contribute to ongoing international debates on the cross-cultural sensitive of the R-SPQ-2F.

4.2 Paper III: Development and validation of calculus self-efficacy inventory

Paper III: Zakariya, Y. F., Goodchild, S., Bjørkestøl, K., & Nilsen, H. K. (2019). Calculus self-efficacy inventory: Its development and relationship with approaches to learning. *Education Sciences, 9*(3), 170. doi:10.3390/educsci9030170

4.2.1 Research aim

The purpose of the study that was reported in Paper III was to develop a calculus self-efficacy inventory with high psychometric properties such as validity and reliability. I suppose that such a task-specific measure of perceived self-efficacy will go a long way in ensuring valid estimates of effect weights in the structural equation analysis of the hypothesised model that was presented in Figure 2.

4.2.2 Research method

I subjected the initial 15-item CSEI to exploratory factor analysis to investigate the factor structure of the CSEI. The exploratory factor analysis provides evidence for the construct validity of the measure. I investigated the discriminant or predictive validity of the measure using Spearman' rank coefficient that accounts for the

[8] The statement of the deleted item is available in Chapter 5.

ordinal scale of measurement of the CSEI. Further, I provided evidence for the reliability index using the latent variable approach.

4.2.3 Main results

The results of the exploratory factor analysis revealed a one-factor structure for the CSEI with minimum rank factor analysis for factor extraction, oblique promin rotation, and parallel analysis procedure for retaining the extracted factors. The one-factor solution was achieved after deleting two items from the initial 15-item CSEI as recommended by the 95 percentiles of the parallel analysis procedure. The results confirm that the measure is unidimensional, i.e., all its items expose a single construct hypothesised to be calculus perceived self-efficacy. The reliability index of the final 13-item CSEI was found to be .90. The positive and negative Spearman' rank coefficients that were found between scores of students on the CSEI and the deep and the surface approaches to learning, respectively, constitute evidence for discriminant or predictive validity of the CSEI. I claim that the findings of this study constitute an original contribution to the literature on the measures of mathematics perceived self-efficacy.

4.3 Paper IV and Paper V: Validation of a test of prior mathematics knowledge

Paper IV: Zakariya, Y. F., Nilsen, H. K., Goodchild, S., & Bjørkestøl, K. (2020). Assessing first-year engineering students' pre-university mathematics knowledge: Preliminary validity results based on an item response theory model. *Journal of Technology and Science Education*, *10*(2), 259-270. doi:10.3926/jotse.1017

Paper V: Zakariya, Y. F., Nilsen, H. K., Bjørkestøl, K., & Goodchild, S. (2020). Impact of attitude on approaches to learning mathematics: a structural equation modeling approach. In T. Hausberger, M. Bosch & F. Chelloughi (Eds.), *Proceedings of the Third Conference of the International Network for Didactic Research in University Mathematics (INDRUM 2020, 12-19 September 2020)* (pp. 268 - 277). Bizerte, Tunisia: University of Carthage and INDRUM.

4.3.1 Research aim

The studies reported in both Paper IV and Paper V are aimed at providing empirical evidence for item quality, construct validity, and reliability of the NMT. The studies attempt to address some issues that surround the following questions:

1. How difficult are the items on the test for the students, and why are they difficult? (item difficulty analysis)

2. Do items on the test discriminate appropriately between students with low and high prior mathematics knowledge that the test was designed to expose? (item discrimination analysis)

3. Does the variance of a single construct responsible for the covariance of students' scores on the items of the test? (unidimensionality aspect of construct validity)

4. How consistent are the items of the test in exposing the construct they are hypothesised to expose? (item reliability analysis)

5. Does the attitude towards mathematics subscale of the NMT possess construct validity, predictive validity, and reliability?

4.3.2 Research method

I applied a two-parameter item response theory model coupled with the latent variable approach for reliability to address questions 1 to 5 that are presented in Section 4.3.1. Further, some students were interviewed to ascertain likely reasons why they perceived some questions on the test to be difficult. Finally, structural equation modelling was used to investigate the predictive validity of the NMT.

4.3.3 Main results

The results of the item response theory analysis confirm that some items of the NMT are too difficult for the students and some items lack appropriate discriminating indices. As an immediate implication of these findings, the items of poor quality (5 of them) were excluded from the test before the final analysis of the main study data. More so, it was revealed that the NMT is unidimensional and its items are reliable coupled with a reliability index of .92 on the whole test. These findings are well-documented in Paper IV. Further analysis in Paper V revealed that the attitudes toward mathematics subscale of the NMT has appropriate construct and a reliability index of .78. The attitudes toward mathematics had a positive effect on deep approaches to learning mathematics and a negative effect on the surface approaches to learning mathematics. These findings constitute evidence of predictive or discriminant validity of the subscale.

4.4 Paper VI: Addressing the research question two

Paper VI: Zakariya, Y. F., Nilsen, H. K., Goodchild, S., & Bjørkestøl, K. (2020). Self-efficacy and approaches to learning mathematics among engineering students: Empirical evidence for potential causal relations. *International Journal of Mathematical Education in Science and Technology*, 1-15. doi: 10.1080/0020739X.2020.1783006

4.4.1 Research aim

The purpose of the study reported in Paper VI was to investigate the potential causal effect of self-efficacy on approaches to learning to mathematics. Therein, I addressed the research question two[9] and tested Hypothesis 5[10] of the present study.

4.4.2 Research method

In this study, I analysed the data at the instance two of phase two data collection, which involved both R-SPQ-2F and CSEI. I evaluated the hypothesised causal relationship between self-efficacy and approaches to learning mathematics using structural equation modelling with WLSMV estimator. This evaluation followed a two-stage measurement-then-structural model analytic procedure as it is typical in the literature (e.g., Byrne, 2012).

4.4.3 Main results

The results revealed that there is a substantial positive effect ($\beta = .54, p < .001$) of perceived self-efficacy on the deep approaches to learning to mathematics and a substantial negative effect ($\beta = -.47, p < .001$) of the former on the surface approaches to learning mathematics (Figure 6). These results may be interpreted to mean, for a unit standardised metric rise in perceived self-efficacy (e.g., cse + 1) there is a corresponding causal effect of .54 times a unit standardised metric rise on the deep approaches to learning, and a corresponding causal effect of .47 times a unit standardised metric decrease in surface approaches to learning among the

---

[9] Recall that the research question two is: Does perceived self-efficacy influence adoption of either deep or surface approach to learning mathematics among first-year engineering students?

[10] Recall that Hypothesis 5 is: There is a causal effect of perceived self-efficacy on engineering students' ongoing approaches to learning a first-year calculus course.

students. As such, these results confirm the plausibility of Hypothesis 5. More so, the results provide empirical evidence for a claim that perceived self-efficacy does influence the adoption of both deep and surface approaches to learning introductory calculus course among first-year engineering students at the University of Agder. That is, a high sense of perceived self-efficacy tends to induce the adoption of the deep approaches to learning, while a low sense perceived self-efficacy tends to induce the adoption of surface approaches to learning. Thus, the results address the research question two of the present study.

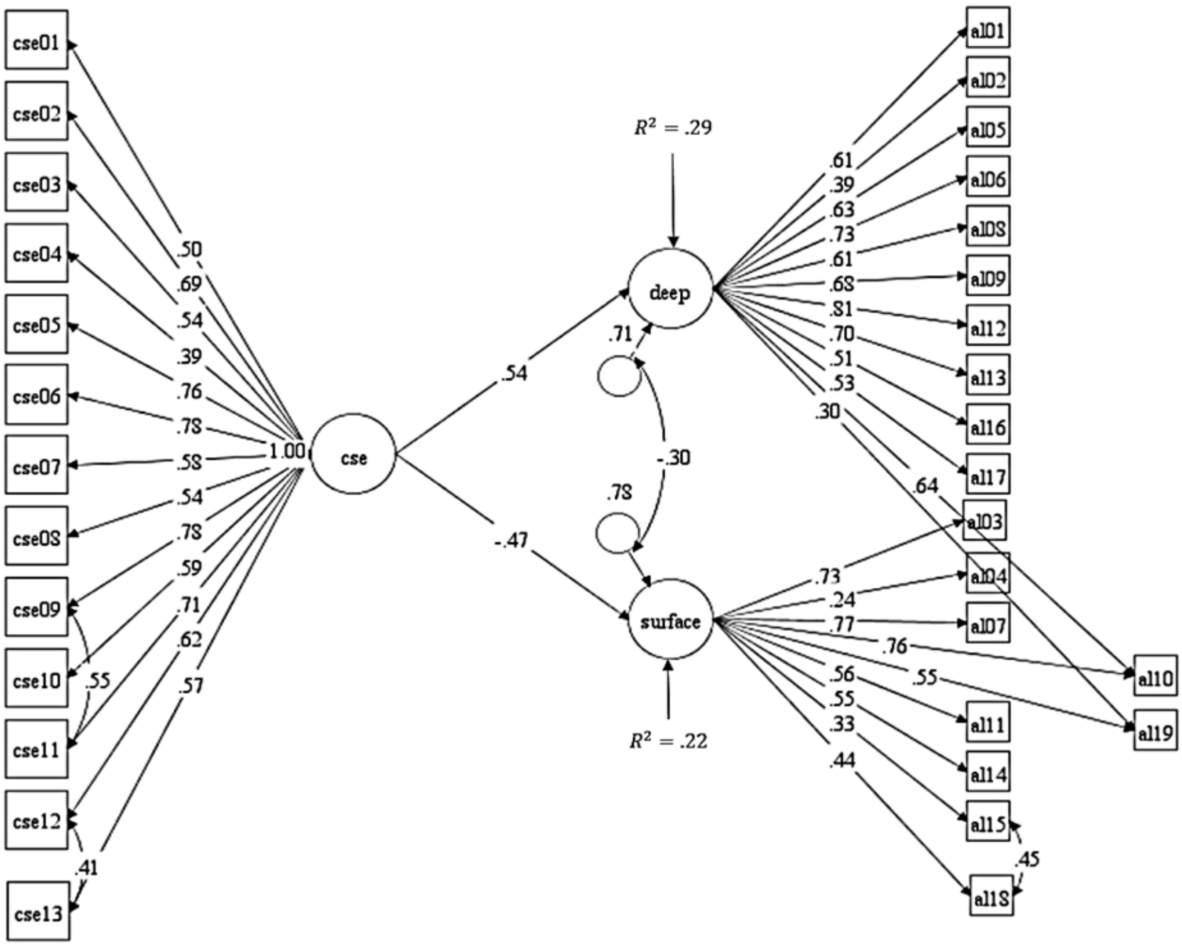


Figure 6. The validated causal model of the hypothesised relationship between self-efficacy and approaches to learning mathematics

Note. The labels ‘cse’: calculus self-efficacy, deep: deep approaches to learning, surface: surface approaches to learning, and R^2: effect size. The labels on the small boxes are item labels of the respective R-SPQ-2F and CSEI. All factor loadings, effect weights, residuals, covariances, and effect sizes are standardised. Figure 6 is reprinted from Paper VI.

4.5 Paper VII: Addressing the research question three (Part A)

Paper VII: Zakariya, Y. F., Nilsen, H. K., Bjørkestøl, K., & Goodchild, S. (forthcoming). Effects of prior mathematics knowledge and approaches to learning on performance in mathematics among first-year engineering students. *European Journal of Education*, under review.

4.5.1 Research aim

The research question three[11] of the present study was split into parts:

> Part A: What are the direct and indirect effects of prior mathematics knowledge and approaches to learning on students' performance in an introductory mathematics course?
>
> Part B: What are the direct and indirect causal effects of prior mathematics knowledge and perceived self-efficacy on performance in mathematics among engineering students?

Part A of the research question three was addressed in Paper VII, and Part B was addressed in Paper VIII. Thus, the study reported in Paper VII aimed to investigate the causal direct effects of both prior mathematics knowledge and approaches to learning mathematics on students' performance in the first-year introductory calculus course. Further, the mediating or indirect effect of prior mathematics knowledge through approaches to learning to mathematics on students' performance in the course was also investigated. By extension, Hypothesis 2[12], Hypothesis 3[13], and Hypothesis 4[14] were evaluated in Paper VII.

---

[11] Recall that the research question three is: What are direct and indirect causal effects of prior mathematics knowledge, approaches to learning, and perceived self-efficacy on performance in mathematics among first-year engineering students?

[12] Recall that Hypothesis 2 is: There is an effect of prior mathematics knowledge on engineering students' ongoing approaches to learning.

[13] Recall that Hypothesis 3 is: There is an effect of engineering students' approaches to learning on their performance in a first-year calculus course.

[14] Recall that hypothesis 4 is: Ongoing approaches to learning mediate the effect of engineering students' prior mathematics knowledge on their performance in a first-year calculus course.

4.5.2 Research method

I identified students' responses to the NMT in instance one of phase two with their responses to the R-SPQ-2F in instance two of phase two data collection and their examination scores. The analyses followed a two-stage measurement-then-structural model analytic procedure as it is typical in the literature (e.g., Byrne, 2012). Confirmatory factor analysis was used to fit the measurement models. After that, I used the structural equation modelling to evaluate the hypothesised causal relationships between the research constructs and to investigate the mediating role of approaches to learning mathematics in the hypothesised model.

4.5.3 Main results

Direct effects

The results revealed that there is a substantial negative effect of prior mathematics knowledge on surface approaches to learning mathematics ($\beta = -.22$, $p < .05$) and a non-significant positive effect of the former on deep approaches to learning mathematics ($\beta = .13$, $p > .05$). One may interpret these findings to mean that students who have low scores on the NMT have a high tendency of adopting surface approaches to learning the calculus course. In contrast, there was no substantial evidence to justify the hypothesised effect of prior mathematics knowledge on the adoption of deep approaches to learning the calculus course. As such, the findings, in parts, confirm the plausibility of Hypothesis 2. More so, surface approaches to learning mathematics have a substantial negative effect on students' performance in the calculus course ($\beta = -.30$, $p < .05$) while the positive effect of deep approaches to learning to mathematics on students' performance in the calculus course was not significant ($\beta < .01$, $p > .05$). It is noteworthy to deduce from these findings that, in as much as, students who adopt surface approaches to learning mathematics performed low in the course there is no substantial evidence to claim that students who adopt deep approaches to learning the course performed better in the course. Thus, Hypothesis 3 is partly confirmed. Furthermore, I found evidence to support the plausibility of the hypothesised effect of prior mathematics knowledge on students' performance in the calculus course (Figure 7). The effect of the former on the latter was significant and positive ($\beta = .20$, $p < .05$).

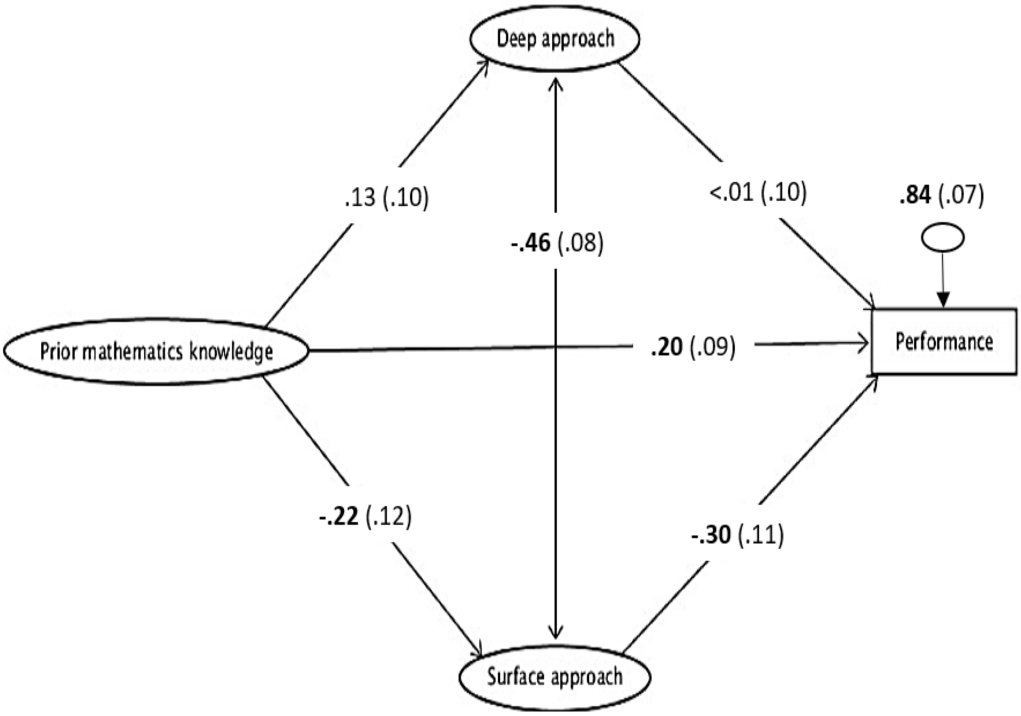


Figure 7. The evaluated structural model of the effects of prior mathematics knowledge and approaches to learning on students' performance in an introductory calculus course

Note. All the estimates are standardised and the significant effect paths at p<.05 are in boldfaces. The estimates in brackets are the corresponding standard errors associated with effect weights. A full figure that contains all the items and the associated parameter estimates is available in the appendix of Paper VII. Figure 7 is reprinted from the paper.

Mediating effects

Finally, the results of the mediation analysis show that surface approaches to learning mathematics mediate the effect of prior mathematics knowledge on students' performance in the course while the deep approaches to learning do not mediate this effect. That is, there is only a substantial indirect effect of prior mathematics knowledge through the surface approaches to learning on students' performance in the course. Thus, evidence suggests that Hypothesis 4 is only plausible for the surface approaches to learning mathematics. Both the direct and the indirect effects that are reported in Paper VII seem to address part A of the research question three of the present study.

4.6 Paper VIII: Addressing the research question three (Part B)

Paper VIII: Zakariya, Y. F. (2021). Self-efficacy between previous and current mathematics performance of undergraduate students: an instrumental variable approach to exposing a causal relationship. *Frontiers in Psychology*. 11:556607. doi:10.3389/fpsyg.2020.556607.

4.6.1 Research aim

The purpose of the study reported in Paper VIII was to investigate the causal effect of perceived self-efficacy on students' performance in a first-year introductory calculus course. Therein, attempts are made to address part B of research question three and to evaluate Hypothesis 6[15], Hypothesis 7[16], and Hypothesis 8[17] of the present study.

4.6.2 Research method

I identified students' responses to the NMT in instance one of phase two with their corresponding responses to the CSE in instance two of phase two data collection and their examination scores. In addition to the test questions of NMT, as a measure of prior mathematics knowledge, I used students' responses to a preliminary item of the NMT on grade points (HGP) in latest pre-university mathematics course they have followed. Confirmatory factor analysis was used to fit the measurement models. After that, I used the structural equation modelling to evaluate the hypothesised causal relationships between the research constructs and to investigate the mediating role of perceived self-efficacy in the hypothesised model. Further, I used an innovative instrumental variable approach to modelling with NMT as an instrumental variable to account for endogeneity problems, as highlighted in Section 3.6.3. As such, I was able to discern the causal effect of perceived self-efficacy on students' performance in the course from other confounding effects of omitted variables in the hypothesised model. According to

---

[15] Recall that Hypothesis 6 is: There is an effect of perceived self-efficacy on engineering students' performance in a first-year calculus course.

[16] Recall that Hypothesis 7 is: There is an effect of prior mathematics knowledge on perceived self-efficacy among first-year engineering students.

[17] Recall that Hypothesis 8 is: Perceived self-efficacy mediates the effect of engineering students' prior mathematics knowledge on their performance in a first-year calculus course.

the literature (e.g., Antonakis et al., 2010; Greenland, 2000), the techniques of the instrumental variable approach to modelling require (a) allowing disturbances of both the perceived self-efficacy and the performance to correlate (Figure 8) such that any common cause of the variables can be captured in the model; (b) introducing an instrumental variable (NMT) in the model that satisfies the following properties: (i) there is a substantial effect of NMT on perceived self-efficacy, (ii) there is a trivial effect of NMT on performance in the model, and (iii) the disturbances of both NMT and the performance are not correlated. It has been statistically shown that the introduction of the instrumental variable in the model allows for an unbiased estimate of the causal effect of perceived self-efficacy on performance in the course (Mulaik, 2009).

4.6.3 Main results

The results (Figure 8) revealed a substantial positive effect (β=.43, $p = .02$) of perceived self-efficacy on students' performance in the calculus course. Thus, confirming the plausibility of Hypothesis 6. It was found that prior mathematics knowledge as exposed by the NMT has a substantial influence on perceived self-efficacy in the course (β=.52, $p < .001$). That is, students with high scores on the prior mathematics knowledge test are ascribed to high sense of perceived self-efficacy on the calculus task. As such, this finding provides empirical evidence for the plausibility of Hypothesis 7. The results of the mediation analysis show that perceived self-efficacy plays a significant mediating role between prior mathematics knowledge and students' performance in the calculus, which confirms Hypothesis 8. Thus, I argue that it is prudent to develop interventions that foster perceived self-efficacy as proxies to enhance students' performance in the course.

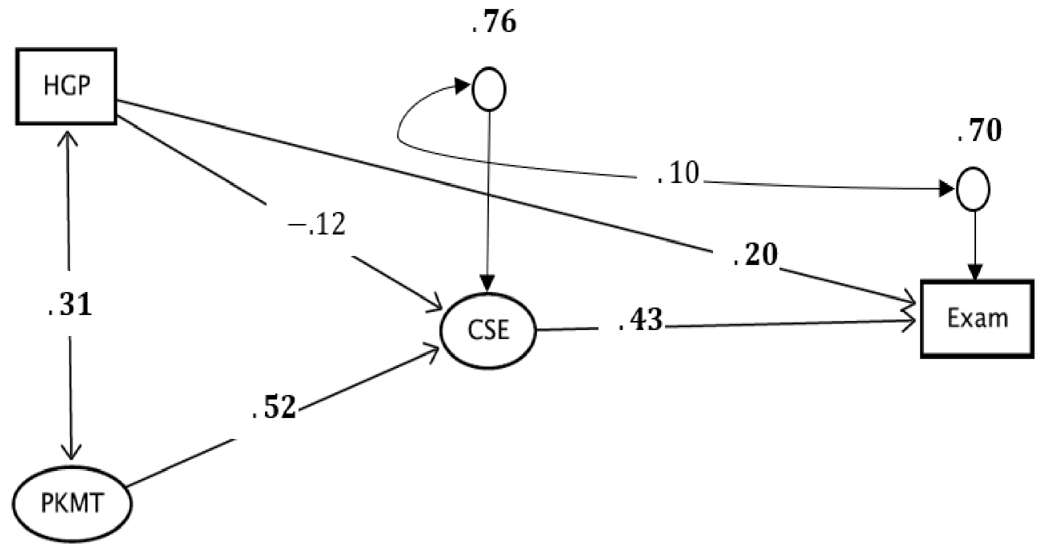


Figure 8. The evaluated model of the relationship between prior mathematics knowledge, self-efficacy, and students' performance in an introductory calculus course

Note. Figure 8 is reprinted from Paper VIII. Both HGP and PKMT (used in Paper VIII in place of NMT in the present dissertation) are measures of the prior mathematics knowledge of the students, CSE is a measure of the self-efficacy, and Exam represents students' performance in the calculus course. The significant estimates and paths are in bold faces, and the items of both PKMT and CSE are not included in Figure 8 to enhance the readability of the figure. The full figure that contains all the items and the associated model parameters is available in the Appendix of Paper VIII.

4.7 Summary of the chapter

The present chapter gives the summaries of published and forthcoming papers that are combined to form the present dissertation. Therein, I highlighted the research aim, peculiar aspects of the research method, and crucial results in each paper. Further, I linked the findings of each paper to the research questions and or research hypotheses they are purported to address. These findings will be discussed in the next chapter.

5 Discussion of, reflection on, and limitations of findings

5.1 Discussion of major findings

In this chapter, I will set out more explanations and interpretations of the main results that are presented in Chapter Four of the present dissertation. However, a complete exposition of this discussion as it relates to the findings of each paper is available in the respective papers. Here, I will highlight some important interpretations and implications of these findings under two broad categories: test validation studies and structural validation studies.

5.1.1 Findings of test validation studies

There appears to be a consensus among quantitative education researchers on the fact that the quality and plausibility of findings resulting from nonexperimental quantitative research depend largely on test validity and reliability. Several pieces of empirical evidence have been provided to support the argument for ensuring test validity and reliability before implementation of such measures in quantitative research (e.g., Bisson, Gilmore, Inglis, & Jones, 2016; Jones, Bisson, Gilmore, & Inglis, 2019; Zakariya, 2020). It does follow from logical reasoning to at least be sure that a measure, indeed, exposes the construct it is purported to expose. Findings of the test validity studies in the present study underscore this logic of reasoning. For instance, contrary to the proposed first-order four-factor and two-factor item-parcelled R-SPQ-2F measurement models by Biggs et al. (2001) I only found evidence to support a two-factor model without item-parcelling for the measure. This evidence was achieved after deletion of one item from the original 20-item measure. The statement of the deleted item is "I learn some things by rote, going over and over them until I know them by heart even if I do not understand them" (Biggs et al., 2001, p. 148). It has been argued that the deleted item is not only statistically poor but also conceptually cumbersome for students (Zakariya, Bjørkestøl, et al., 2020). More so, the finding conforms to a body of literature where researchers have recommended deletion of some items from the original R-SPQ-2F to achieve sufficient evidence of its construct validity (e.g., Immekus & Imbrie, 2010; López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014; Stes, De Maeyer, & Van Petegem, 2013).

More importantly, the findings reported in Paper I and Paper II confirm that deep and surface approaches to learning are prevalent among engineering students. Thus, the findings appear to address the first research question and Hypothesis 1 of the present study. I would like to emphasise that this prevalence of deep and surface approaches is a characterisation of students' learning processes at the group level. As such, there could be a discrepancy in terms of identifying an individual student with either deep or surface approaches to learning mathematics. Even though this characterisation of each student into either deep or surface approaches to learning does not have any substantial impact on the present study, yet I may recommend a qualitative research methodology to explore this possibility.

The validity and reliability evidence I exposed in the present study for a measure of calculus self-efficacy is noteworthy. The findings confirm my confidence in the measure and are important to the validity of results from the structural evaluation of the hypothesised model in the main study. Similarly, my confidence is also reinstated by the findings reported in Paper IV and Paper V on the measure of prior mathematics knowledge. Previous studies are lacking in mathematics education literature to compare my findings on the psychometric properties of the self-efficacy inventory as well as the NMT. However, there are some notable advantages of CSEI such as its short length, its high predictive validity, and its reliability index over similar measures of mathematics self-efficacy that have been reported, elsewhere (e.g., Betz & Hackett, 1983; Kranzler & Pajares, 1997).

It is important to acknowledge that the issues of test validity and reliability are a bit complex and are subject to open questions among methodologists and researchers especially when it comes to test score meanings and interpretations (Kane, 2013; Markus & Borsboom, 2013). As succinctly put by Kane (2012),

> Validity [in a broad sense] raises some difficult questions, and it would be unreasonable to expect that the answers will be simple or formulaic. Some aspects of validity tend to get quite technical (e.g. predictive models, statistical models of bias); some are more philosophical (e.g. causal inferences); and some raise broad social issues (fairness, intended and unintended consequences) (p. 4).

I do not claim to have exposed evidence of test validity for all the measures used in the present study in the broad sense of validity. I firmly believe that to achieve such an integrative perspective of validity for a measure requires more than one study and, indeed, multiple evidence is needed. However, I concentrate only on an aspect of test validity within the latent theory of measurement by making, evaluating, and deducing causal claims between theorised constructs and observed variables to provide validity evidence for the measures of research constructs. The data collected and analysed provide an empirical basis for the plausibility of the hypothesised relations. More studies are required to strengthen these relationships in independent samples. It is still an open question whether the students understand or make meaning of the items of the measures that I developed or validated with the intended understanding or meaning by the researcher. Future studies are recommended to explore this open question.

5.1.2 Findings of structural validation studies

The research question two

The first crucial finding after the validations of research measures addresses the second research question of the present study. The finding shows that there is a substantial influence of perceived self-efficacy on students' approaches to learning mathematics with a positive influence on the deep approaches and a negative influence on the surface approaches to learning. This finding, on the one hand, means that a high sense of perceived self-efficacy tend to induce the adoption of a deep approach to learning mathematics. That is, students with a high sense of perceived self-efficacy tend to:

- Be excited about new mathematics topics and devote their spare time to develop a proper understanding of the topics.
- Self-test themselves on crucial mathematics topics to develop mastery of the subject matter.
- Study hard for mathematics because of personal interest and feeling of satisfaction in the course.
- Be highly prepared for mathematics classes with unanswered questions during their self-study prior to their class attendance.

- Explore suggested readings for the course to develop more calculation skills.

On the other hand, the finding means that a low sense of perceived self-efficacy tend to induce adoption of surface approaches to learning mathematics. That is, students with a low sense of perceived self-efficacy tend to:

- Aim only at passing the course with limited work done.
- Use memorisation techniques often with less care for developing a proper understanding of the mathematics content.
- Think that remembering answers to plausible examination questions is the best method to pass the examinations.
- Think that in-depth preparation for classes or study of mathematics topics is unnecessary; it wastes time and confusing.
- Self-confine themselves to class materials with a thought that is unnecessary to solve extra mathematical tasks.

The confirmed effect of perceived self-efficacy on approaches to learning mathematics has a potential implication to engineering course coordinators, university teachers, and other education stakeholders such as MatRIC who are directly involved in the teaching of mathematics to engineering students. This implication is the provision of evidence to support the logic of designing perceived self-efficacy interventions as proxies to influence the adoption of students' approaches to learning the course. Further, it can be argued that the effect of perceived self-efficacy on approaches to learning mathematics complements previous studies in science, technology, engineering, and mathematics (STEM) on the relationships between the constructs (e.g., Ardura & Galán, 2019; Shen, Lee, Tsai, & Chang, 2016). Complementary in the sense that the finding exposes the relationship between the constructs within the mathematics education literature which is lacking in the STEM research.

The research question three

The second set of crucial findings after the validations of research measures addresses the third research question of the present study. These findings expose the direct and indirect causal effects of prior mathematics knowledge, approaches to learning mathematics, and perceived self-efficacy on performance in mathematics among first-year engineering students. Of prime importance among these findings is the substantial influence of surface approaches to learning mathematics on students' performance in the course while deep approaches to learning mathematics fail to exert the expected influence on students' performance in the course. An immediate implication of these findings to calculus teachers may be a need to discourage the students (during classroom sessions, consultations, and drop-in sessions) from using the following strategies while studying for the course:

- Aiming only at passing the course with limited work done.
- Use memorisation techniques often with less care for developing a proper understanding of the mathematics content.
- Thinking that remembering answers to plausible examination questions is the best method to pass the examinations.
- Thinking that in-depth preparation for classes or study of mathematics topics is unnecessary, it wastes time and confusing.
- Self-confining themselves to class materials with a thought that is unnecessary to solve extra mathematical tasks.

These strategies are not the best. In fact, evidence in the present study and elsewhere (e.g., Mundia & Metussin, 2019; Nguyen, 2016) shows that they may lead to low performance in the course. Therefore, it is not surprising that surface approaches to learning mathematics are the only mediators of the effect of prior mathematics knowledge on students' performance in the course. This is because the effect of the deep approaches to learning mathematics on the students' performance is not significant. The fact that the deep approaches to learning mathematics do not influence students' performance in the course, contrary to my expectation, has antecedents in the literature (e.g., García et al., 2016). As such, one may ascribe this pattern of relationship between deep approaches to learning mathematics and students' performance in the course to the nature of course assessment. Perhaps, the end of semester examination favours assessment of procedural knowledge rather than conceptual knowledge. In which case, deep approaches to learning mathematics that are presumed to lead to better conceptual understanding are not assessed. Even though some researchers (e.g., Rittle-

Johnson & Alibali, 1999) have shown that both the procedural and conceptual knowledge of mathematics are causally related and as such the distinction between the types of understanding is faint.

Another crucial finding in the present study that addresses some aspects of the research question three is the substantial causal effect of perceived self-efficacy on students' performance in the course. This finding means that perceived self-efficacy is a potential cause of students' performance in the course. That is, students with a high sense of perceived self-efficacy performed better in the course than the students with a low sense perceived self-efficacy and that a cause of this better performance may be linked to the variability in the students' perceived self-efficacy. I claim that this finding is unique from two perspectives. First, the application of the instrumental variable approach to discern the causal effect of perceived self-efficacy from that of other omitted confounding effects on students' performance in the course is unprecedented. To the best of my knowledge, I am not aware of any quantitative research in mathematics education that has applied this innovative approach to expose the causal effect of perceived self-efficacy on undergraduate students' performance in mathematics. Second, the exposed causal effect of perceived self-efficacy on students' performance in the course complements previous research within mathematics education (e.g., Pajares & Kranzler, 1995; Pajares & Miller, 1994). Complementary in the sense that the finding provides state of the art evidence on the causal relationship between the constructs.

Furthermore, perceived self-efficacy was revealed to substantially mediate the effect of prior mathematics knowledge on students' performance in the course. At this juncture, I think it is necessary to highlight one significance of this finding. Recall that prior mathematics knowledge has earlier been reported in Paper VII to have a substantial effect on students' performance in the course. However, when perceived self-efficacy was included in the structural model, it absorbed, almost wholly, this effect of prior mathematics knowledge on students' performance in the course. Moreover, it has earlier been reported that perceived self-efficacy influences adoption of approaches to learning mathematics (Zakariya, Nilsen, Goodchild, & Bjørkestøl, 2020b). Thus, as an implication of these findings, with respect to the studied constructs, one can argue that it seems prudent to develop or implement existing interventions that foster perceived self-efficacy with a focus on calculus tasks. If such interventions are implemented, it appears plausible that improved learning outcomes in the course will be achieved among the engineering students. The next question should be what are the available evidence based-based interventions that foster perceived self-efficacy on mathematics tasks? I will attempt to address this question in the next chapter.

5.2 Reflection on the findings

In Section 5.2, I have attempted to highlight some essential interpretations and implications of findings that emanated from the eight papers that make up the present dissertation. It is crucial to acknowledge that each of these papers has specific 'local' aims and objectives (Section 4.1.1, Section 4.2.1, …, and Section 4.6.1) toward achieving the 'global' aims of the present dissertation (Section 1.3). However, I believe some findings of these studies are prerequisites for other studies while some findings are complementary to each other toward making a coherence argument for the dissertation. For instance, the validity and reliability evidence of the R-SPQ-2F that are presented in Paper I and Paper II are prerequisites for the studies reported in Paper VI and Paper VII to ensure reliable estimates of the causal relationships that are reported in these papers. On the flip side, the findings that are reported in Paper VI, Paper VII and Paper VII are bits that strengthen each other to make a whole toward addressing the research questions two and three of the present study.

The findings of the present study have shown that only surface approaches to learning to mathematics have a substantial negative influence on students' performance in the introductory calculus course. In which case, the positive influence of deep approaches to learning mathematics on students' performance in the course is not significant. Even though, these findings are contrary to the findings of some researchers (e.g., Maciejewski & Merchant, 2016; Mundia & Metussin, 2019) the findings conform to the report by Nguyen (2016). Thus, the present study seems to provide a clear understanding on the specific type of approaches to learning mathematics that influences first-year students' performance in an introductory calculus course within the target population of students of the present study. I believe that these findings constitute an attempt to fill the related knowledge gap as exposed in Section 1.2.1. It is important to remark that the present study is confined to only two, supposedly, distinct types of approaches to learning to mathematics (deep and surface approaches) in which no attempt is made to investigate a hybrid type or other types of approaches to learning the course. I envisage that future studies will be conducted with this intention.

Further, the findings of the present study provide state-of-the-art evidence on the causal effect of perceived self-efficacy on students' performance in a first-year introductory calculus course. Given that the structural equation modelling technique used to expose this causal effect follows an item-level modelling instead of composite score modelling (e.g., Pajares & Kranzler, 1995) I claim that the exposed causal effect has high accuracy in the representation of the reality (Bandalos, 2008). In addition, the instrumental variable approach to modelling used in exposing this causal effect strengthen the estimate of a causal relationship between the constructs. As such, the related knowledge gap as exposed in Section 1.2.2 seems to be addressed. However, I acknowledge that there are some unanswered questions in the present study and the causal inferences remain controversial, especially from a nonexperimental study (Freedman, 2004). On this note, I point out to the following statements by McDonald (2011):

> Perhaps enough has been said to point up the difficulties attending causal inference from nonexperimental data. We must also face the fact that if investigators took the view that such inferences should never be attempted, many of the most important questions in the behavioural and social sciences would remain entirely out of the bounds of ethical and otherwise practical modes of research. (p. 371)

The findings of the present study also provide empirical evidence for indirect effects of prior mathematics knowledge through both approaches to learning mathematics and perceived self-efficacy to students' performance in the course. These findings constitute an attempt to fill the related knowledge gap that is exposed in Section 1.2.3. It would have been more interesting if the indirect effects of prior mathematics knowledge have been investigated with a complex model, as presented in Figure 2, rather than splitting the model into two and investigating them separately as reported in Paper VII and Paper VIII. Nevertheless, I could not investigate the complex model as proposed in Figure 2 because of the small resultant sample (less than 90 respondents) when data from instance one and instance two of the main study data collection were combined on all the variables. More so, some statistical techniques for handling missing data in SEM (e.g., full information maximum likelihood) could not help the situation given the categorical level of measurement of most the research variables. As Mulaik (2009) rightly put it “It is unrealistic to suppose that one will always be able to perform a study with an SEM and get everything right the first time. Science progresses in graduated steps with series of studies” (p. 188). Therefore, all the findings of the present study provide tentative evidence for the established relationships between the research constructs that are subject to future validations.

5.3 Limitations and generalisation of findings

Despite the contributions of the present study to the literature and communication of its findings to agents of implementations[18], there are some limitations that are worth mentioning. Some of these limitations are from the conceptualisation of this study, while several others are from the methodology and methods. For instance, my choice of concentrating on student-source factors has undoubtedly cut out some crucial factors that affect students’ performance in the course, e.g., collective dimensions and the students’ activity systems, students’ learning experience, teacher-student interaction, students’ mathematical discourse, contradictions, and tensions in learning first-year calculus course. Even within the student-source factors, my argument in Chapter One for the superiority of approaches to learning mathematics and perceived self-efficacy does not cover all the student-source factors. As such, I must admit that the findings of the present study are confined to the research constructs. However, I think it is not expected to proffer absolute solutions to students’ poor performance in a single project given that the problem of poor performance is multi-dimensional. Instead, multiple studies on different dimensions of the problem are expected. Thus, I recommend more studies on other factors that affect students’ performance in mathematics to complement the findings of the present study.

The confinement of the research sample to first-year engineering students may constitute a limitation to the findings of the present study. Even though, there are obvious advantages and cogent reasons for focusing on this set of students as earlier highlighted in Chapter One and Chapter Three. It is logical to argue that more interesting findings would have been exposed had the present study covered a student’s population with greater diversity. On this basis, I recommend future

[18] An exposition on the research outreach is presented in the next chapter.

research on applications of the framework and analytic tools that are developed in the present study to different and diverse sets of students. It is also important to mention that it would have been more interesting if the innovative instrumental variable approach to modelling was used in Paper VII to expunge the confounding effects due to omitted variables on students' performance in the course. As such, more refined causal effects of approaches to learning mathematics on students' performance in the course would have been exposed. However, I was not able to use the instrumental variable approach to modelling in Paper VII due to lack of a suitable *instrument* for the analysis. I suppose future research would explore this possibility.

It is acknowledged that some of the highlighted limitations of the present study pose challenges to the generalisation of the findings. Meanwhile, the relatively large samples of the present study coupled with the type of statistical tools used to evaluate the theory-driven hypothesised relationships between the research constructs offer some confidence to the generalisation of the findings. I am confident that the psychometric properties of the NMT that are reported in the present study are generalisable to the national context. Notwithstanding, I advise the National Mathematical Council to replicate my study with a larger sample. Similarly, the findings from the validation studies of both the R-SPQ-2F and the CSEI are generalisable to first-year university students that followed the introductory calculus course in Norway. More importantly, I conjecture that similar patterns of relationships between prior mathematics knowledge, approaches to learning mathematics, perceived self-efficacy, and students' performance in first-year mathematics will be exposed if the presented study is replicated beyond the Norwegian borders.

5.4 Summary of the chapter

I have attempted to shed more light on the crucial findings of the present study. Therein, I related the findings from both the pilot and the main studies to what is known in related literature. I exposed some implications of the findings to agents of implementation. I registered my dispositions toward the concept of test validity, the intended validity evidence exposed in papers I-V, and the necessity for more validation studies to strengthen my findings. Further, I reflected on the contributions of each paper toward making a coherence argument for the achievement of the research aims. Finally, I acknowledged some potential limitations of this study and the implications of these limitations to the generalisation of findings.

6 Research outreach, significance of findings, and conclusion

6.1 Research outreach and significance of findings

I believe that the utility of education research is not effectively actualised until the findings of such research are communicated to agents of implementation. These agents of implementation may be classroom teachers from whom students' data are collected, course coordinators, curriculum planners, research centres, e.g., MatRIC, research councils, e.g., National Mathematical Council in Norway, and the education research community. Many researchers concentrate on reaching out to the education research community through presentations of their research findings at conferences, seminars, workshops, and through journal article publications. In doing so, they tend to give less attention to other agents of implementation. I take to a more holistic approach of communicating the research findings of the present study by directly addressing the classroom teachers of mathematics at the university, a research centre, and the National Mathematical Council. For instance, I attended a seminar organised by MatRIC for PhD fellows and engineering mathematics teachers on May 13, 2019. Therein, I presented the initial conceptualisation of my project and received feedback from international mathematics education researchers that attended the conference from the United States, United Kingdom, and Sweden.

More so, I engaged the National Mathematical Council in Norway on crucial findings of my validation studies on the NMT. I communicated to the Council about specific item quality of the test such as item discrimination and difficulty indices as well as predictive validity and reliability of the test. I made some recommendations for possible improvement of the test and its scoring procedure, and I offered an option to share full papers detailing with the analysis and critical interpretation of the findings in any forum they believe appropriate. To the best of my knowledge, this overarching test validation study on NMT is unprecedented within the Norwegian borders. As such, the significance of my findings may be perceived in shaping the future administration of the test. Thus, I claim that the present project is making a national impact through this research outreach. The validation of the R-SPQ-2F makes a substantial contribution, at least from a Norwegian context perspective, to the ongoing debate on the cultural sensitivity of the measure (e.g., Whitelock-Wainwright, Gasevi, Wood, & Ryan, preprint).

Similarly, the development and validation of the calculus self-efficacy inventory constitute an original contribution to the literature.

Furthermore, I engaged the classroom teachers from whom students' data were collected, engineering course coordinators, curriculum planners including the dean of faculty of engineering at the University of Agder, MatRIC leaders to discuss crucial findings of my project. In the seminar, I presented the significant findings of the present study from the research question one to the research question three. I highlighted the substantial influence of perceived self-efficacy on students' performance in the first-year introductory calculus course as revealed from my study. After that, there was a general discussion on a matter arising from the presentation, including potential interventions that may foster students' perceived self-efficacy. I believe, through this seminar, that I did send signals to appropriate agents of implementation on potential solutions in alleviating students' poor performance in the course. Admittedly, to improve students' performance in a course requires a holistic approach to the problem. However, I am confident that the findings of the present study have exposed the areas to look at when it concerns student-source factors in solving this problem. It is my opinion that the publications of findings of the present study in different peer-reviewed internationally recognised journals and a conference are sufficient in reaching out to education research community that is involved in the teaching and learning of mathematics at higher education. It is expected that these findings will reach a broad community of education researchers because all the papers are published as open access.

6.2 Concluding remarks

The present study was motivated by the poor performance of engineering students in a first-year introductory calculus course at the University of Agder. Therein, attempts were made to provide empirical evidence on the areas of concentration, as it concerns the student-source factors, to alleviate this problem. Previous studies suggest that prior mathematics knowledge, approaches to learning mathematics, and perceived self-efficacy, among other student-source factors, play the most significant roles in fostering students' performance in mathematics. However, most of these previous studies are correlational (e.g., traditional regression-based studies), which makes it difficult to argue for the effectiveness of interventions on these factors as proxies to enhance performance in the course. This is because a logical inference of the effect of one variable on the other after implementation of interventions requires causal assumptions between the variables (Pearl & Mackenzie, 2019). Thus, the present study was framed within the quantitative research paradigm and data were collected using mainly questionnaires and tests. The collected data were analysed using causal modelling techniques such that causal claims are made and evaluated.

The evidence from a series of structural equation modelling techniques points to the fact that perceived self-efficacy (engineering students' convictions to solve first-year introductory calculus tasks successfully) has the most substantial effect on the students' performance in the course. Its effect overshadows the effects of both prior mathematics knowledge and approaches to learning mathematics on students' performance in the course. Further, perceived self-efficacy appears to influence adoption of either deep or surface approaches to learning mathematics among the engineering students. Therefore, a major conclusion drawn from the findings of the present study is the identification of perceived self-efficacy as prime factor whose interventions could enhance students' performance in the course. As such, I conclude the present dissertation by highlighting two evidence based-based interventions that foster students' perceived self-efficacy in mathematics. These interventions will serve as potential cues in solving the problem of poor performance in mathematics if implemented by the university teachers, education researchers, policymakers, and other stakeholders who are involved in teaching and learning of first-year undergraduate mathematics courses.

6.3 Evidence-based self-efficacy interventions

Recall that there are four sources of perceived self-efficacy: "enactive mastery experience", "vicarious experience", "verbal persuasions", and "physiological and affective states" (Bandura, 1997, p. 79), as highlighted in Chapter two. It is prudent, and of course, the approach adopted in perceived self-efficacy intervention studies, to intervene through these sources for improved perceived self-efficacy of the students. Two of these studies are summarised in the following sections, including some ideas about possible implementations in the teaching and learning of introductory calculus course at the University of Agder. One can build on these examples to develop and implement similar interventions in further studies.

6.3.1 Mathematical modelling competition

Evidence has shown that mathematical modelling competitions have the potential to foster perceived self-efficacy on mathematics tasks. Czocher et al. (2019) demonstrated how an extra-curricular modelling competition led to post-secondary school students' gains in perceived self-efficacy with a focus on applications of differential equations. The basic idea of this competition is to design modelling tasks drawn from the syllabus of a target course (in our case, calculus course) and invite students to participate outside the regular classroom teaching. Several examples of modelling tasks based on ideas from introductory calculus course are available, elsewhere, (e.g., Kilty & McAllister, 2018). The students can be divided into groups with each group voluntarily choosing a modelling task to work on during the competition. The students are expected to work collaboratively to solve the chosen problem using concepts and methods from the calculus course, seek mentorship from the teachers or older colleagues, and make a presentation on the day of the competition. The inherent mechanisms that enhance perceived self-efficacy through the competition are students' research experience, mentorship from the teachers or older colleagues and community involvement through collaboration with peers (Czocher et al., 2019). It is expected that this extra-curricular activity will provide students' with mastery experience on the content of the course, which is an important source of perceived self-efficacy (Bandura, 2008). As such, I recommend a mathematical modelling competition to foster perceived self-efficacy on calculus tasks among engineering students.

6.3.2 Vicarious experience presentation

A video or live presentation by peers who have passed through a course has been shown empirically to be efficient in fostering students' perceived self-efficacy on the course (e.g., Bartsch, Case, & Meerman, 2012; Luzzo, Hasper, Albert, Bibby, & Martinelli, 1999). The primary idea of the vicarious experience presentation is that the calculus teachers encourage the students to watch short videos of previous students (live models) who have followed and passed the course or invite the live models for a classroom presentation. In the presentation, the previously successful students on the course will narrate their experience, their challenges, their perseverance, their study approaches, and how they manage to pass the course. It has been suggested that a live presentation to provide vicarious experience could be better than a video presentation due to the presence of real-time interaction in the former (Bartsch et al., 2012). Further, previous successful students with an average performance in the course are recommended to give the presentation. It is expected that by watching or monitoring the experience of peers with a similar level of performance will foster the perceived self-efficacy of the current students following the course. The idea is that *if they can, then I can*. Thus, I recommend this intervention to foster perceived self-efficacy on calculus tasks among engineering students.

List of references

Alkhateeb, H. M., & Hammoudi, L. (2006). Attitudes toward and approaches to learning first-year university mathematics. *Perceptual and motor Skills, 103*(1), 115-120. doi:10.2466/pms.103.1.115-120

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: a review and recommendations. *The Leadership Quarterly, 21*(6), 1086-1120. doi:10.1016/j.leaqua.2010.10.010

Ardura, D., & Galán, A. (2019). The interplay of learning approaches and self-efficacy in secondary school students' academic achievement in science. *International Journal of Science Education, 41*(13), 1723-1743. doi:10.1080/09500693.2019.1638981

Asikainen, H., & Gijbels, D. (2017). Do students develop towards more deep approaches to learning during studies? A systematic review on the development of students' deep and surface approaches to learning in higher education. *Educational Psychology Review, 29*(2), 205-234. doi:10.1007/s10648-017-9406-6

Bandalos, D. L. (2008). Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling: A Multidisciplinary Journal, 15*(2), 211-240. doi:10.1080/10705510801922340

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-147. doi:10.1037/0003-066X.37.2.122

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology, 4*(3), 359-373.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148. doi:10.1207/s15326985ep2802_3

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1-26. doi:10.1146/annurev.psych.52.1.1

Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology: An international Review, 51*(2), 269-290. doi:10.1111/1464-0597.00092

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (Vol. 5, pp. 307–337). Greenwich, CT: Information Age Publishing.

Bandura, A. (2008). An agentic perspective on positive psychology. In S. J. Lopez (Ed.), *Praeger perspectives. Positive psychology: Exploring the best in people* (Vol. 1, pp. 167-196). Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*(1), 9-44. doi:10.1177/0149206311410606
Bartsch, R. A., Case, K. A., & Meerman, H. (2012). Increasing academic self-efficacy in statistics with a live vicarious experience presentation. *Teaching of Psychology, 39*(2), 133-136. doi:10.1177/0098628312437699
Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238
Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23,* 329-345.
Biggs, J. B. (1987). *The study process questionnaire (SPQ): Manual*. Hawthorn, Vic.: Australian Council for Educational Research.
Biggs, J. B. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Educational Research Journal, 63,* 3-19. doi:10.1111/j.2044-8279.1993.tb01038.x
Biggs, J. B. (2012). What the student does: Teaching for enhanced learning. *Higher Education Research & Development, 31*(1), 39-55. doi:10.1080/07294360.2012.642839
Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149. doi:10.1348/000709901158433
Biggs, J. B., & Tang, C. (2007). *Teaching for quality learning at university* (3 ed.). Berkshire: The Society for Research into Higher Education & Open University Press.
Bisson, M.-J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 141-164. doi:10.1007/s40753-016-0024-3
Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation. In S. L. Morgan (Ed.), *Handbooks of Sociology and Social Research* (pp. 301-328). Dordrecht: Springer.
Borgonovi, F., & Pokropek, A. (2019). Seeing is believing: Task-exposure specificity and the development of mathematics self-efficacy evaluations. *Journal of Educational Psychology, 111*(2), 268-283. doi:10.1037/edu0000280
Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research & Perspective, 6*(1-2), 25-53. doi:10.1080/15366360802035497
Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203-219. doi:10.1037/0033-295X.110.2.203
Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, London: The Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258. doi:10.1177/0049124192021002005

Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling: A Multidisciplinary Journal, 1*(3), 253-267. doi:10.1080/10705519409539977

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge, Taylor & Francis Group.

Cano, F., Martin, A. J., Ginns, P., & Berbén, A. B. G. (2018). Students' self-worth protection and approaches to learning in higher education: Predictors and consequences. *Higher Education, 76*(1), 163-181. doi:10.1007/s10734-017-0215-0

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science, 3*(3), 186-190. doi:10.1111/j.1467-9280.1992.tb00024.x

Czocher, J. A., Melhuish, K., & Kandasamy, S. S. (2019). Building mathematics self-efficacy of STEM undergraduates through mathematical modelling. *International Journal of Mathematical Education in Science and Technology*, 1-28. doi:10.1080/0020739x.2019.1634223

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment, 23*, 225-241. doi:10.1177/073428290502300303

Dochy, F., De Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning. How far have we progressed since Bloom? Implications for educational practice and teaching. *Active learning in higher education, 3*(3), 265–284. doi:10.1177/1469787402003003006

Dowker, A., Cheriton, O., Horton, R., & Mark, W. (2019). Relationships between attitudes and performance in young children's mathematics. *Educational Studies in Mathematics, 100*(3), 211-230. doi:10.1007/s10649-019-9880-5

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: what have we learned in 60 years? *Frontiers in Psychology, 7,* 508. doi:10.3389/fpsyg.2016.00508

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.

Eisenhart, M. A. (1991). *Conceptual frameworks for research circa 1991: Ideas from a cultural anthropologist; implications for mathematics education researchers.* Paper presented at the 13th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Blacksburg, Virginia, USA.

Eklund, H. (2019). An interactive delivery model for students repeating calculus for scientists & engineers I. *Primus,* 1-19. doi:10.1080/10511970.2018.1506535

Eng, T. H., Li, V. L., & Julaihi, N. H. (2010). The relationships between students' underachievement in mathematics courses and influencing factors. *Procedia - Social and Behavioral Sciences, 8*, 134-141. doi:10.1016/j.sbspro.2010.12.019
Entwistle, N. (2005). Contrasting perspectives on learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 106-125). Edinburgh: The University of Edinburgh, Centre for Teaching, Learning and Assessment.
Entwistle, N., & Tait, H. (1994). *The revised approaches to studying inventory*. Edinburgh: Centre for Research into Learning and Instruction, University of Edinburgh.
Entwistle, N., & Waterston, S. (1988). Approaches to studying and levels of processing in university students. *British Educational Research Journal, 58*, 258-265. doi:10.1111/j.2044-8279.1988.tb00901.x
Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology, 21*, 291-313. doi:10.2307/270939
Freedman, D. A. (2004). Graphical models for causation, and the identification problem. *Evaluation Review, 28*(4), 267-293. doi:10.1177/0193841X04266432
Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*(3). doi:10.7275/n560-j767
García, T., Rodríguez, C., Betts, L., Areces, D., & González-Castro, P. (2016). How affective-motivational variables and approaches to learning predict mathematics achievement in upper elementary levels. *Learning and Individual Differences, 49*, 25-31. doi:10.1016/j.lindif.2016.05.021
Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology, 29*(4), 722-729. doi:10.1093/ije/29.4.722
Gynnild, V., Tyssedal, J., & Lorentzen, L. (2005). Approaches to study and the quality of learning. Some empirical evidence from engineering education. *International Journal of Science and Mathematics Education, 3*, 587–607.
Hackett, G., & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance. *Journal for Research in Mathematics Education, 20*(3), 261-273.
Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: A structural model. *Educational Psychology, 28*(1), 59-71. doi:10.1080/01443410701413753
Hölder, O. (1901). Die Axiome der Quantitat und die Lehre vom Mass. Berichte iiber die Verbandlungen der Koniglich Sacbsischen Gesellschaft der Wissenschajten pu Leippig. *Mathematisch-Physische Klasse, 53*, 1-64.
Hounsell, D. (2005). Understanding teaching and teaching for understanding. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of*

learning: Implications for teaching and studying in higher education (3rd (Internet) ed., pp. 238-257). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.

Hu, L. t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453. doi:10.1037/1082-989X.3.4.424

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Immekus, J. C., & Imbrie, P. K. (2010). A test and cross-validation of the revised two-factor study process questionnaire factor structure among western university students. *Educational and Psychological Measurement, 70*(3), 495-510. doi:10.1177/0013164409355685

Joët, G., Usher, E. L., & Bressoux, P. (2011). Sources of self-efficacy: An investigation of elementary school students in France. *Journal of Educational Psychology, 103*(3), 649-663. doi:10.1037/a0024048

Jones, I., Bisson, M., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal, 45*(3), 662-680. doi:10.1002/berj.3519

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: Guide to the program and applications.* (2nd ed.). Chicago, IL: SPSS.

Kane, M. (2012). Validating score interpretations and uses. *Language Testing, 29*(1), 3-17. doi:10.1177/0265532211417210

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. doi:10.1111/jedm.12000

Kilty, J., & McAllister, A. M. (2018). *Mathematical modeling and applied calculus*. Oxford: Oxford University Press.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Konstantopoulos, S., & Shen, T. (2016). Class size effects on mathematics achievement in Cyprus: Evidence from TIMSS. *Educational Research and Evaluation, 22*(1-2), 86-109. doi:10.1080/13803611.2016.1193030

Kranzler, J. H., & Pajares, F. (1997). An exploratory factor analysis of the mathematics self-efficacy scale revised (MSES-R). *Measurement and Evaluation in Counseling and Development, 29*(4), 215-228.

Kuhn, S. T. (2012). *The structure of scientific revolutions: 50th anniversary edition*. Chicago 60637: The University of Chicago Press.

Leaper, C. (2011). More similarities than differences in contemporary theories of social development?: a plea for theory bridging. In J. B. Benson (Ed.), *Advances in Child Development and Behavior* (2011/09/06 ed., Vol. 40, pp. 337-378). San Diego: Elsevier.

Lester, F. K., Jr. (2010). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. In B. Sriraman & L.

English (Eds.), *Theories of Mathematics Education, Advances in Mathematics Education* (pp. 67-85). Berlin Heidelberg: Springer-Verlag.
Lincoln, Y. S., & Guba, E. G. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (pp. 191-215). London: SAGE.
Lonka, K., Olkinuora, E., & Mäkinen, J. (2004). Aspects and prospects of measuring studying and learning in higher education. *Educational Psychology Review*, *4*(2), 301-323. doi:10.1007/s10648-004-0002-1
López-Aguado, M., & Gutiérrez-Provecho, L. (2018). Checking the underlying structure of R-SPQ-2F using covariance structure analysis / Comprobación de la estructura subyacente del R-SPQ-2F mediante análisis de estructura de covarianza. *Cultura y Educación/ Culture and Education, 30*(1), 105-141. doi:10.1080/11356405.2017.1416787
Luce, R. D., & Suppes, P. (2002). Representational measurement theory. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3 ed., Vol. 4, pp. 1-41). New York: Wiley.
Luzzo, D. A., Hasper, P., Albert, K. A., Bibby, M. A., & Martinelli, E. A., Jr. (1999). Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students. *Journal of Counseling Psychology, 46*(2), 233–243. doi:10.1037/0022-0167.46.2.233
MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149. doi:10.1037/1082-989X.1.2.130
Maciejewski, W., & Merchant, S. (2016). Mathematical tasks, study approaches, and course grades in undergraduate mathematics: a year-by-year analysis. *International Journal of Mathematical Education in Science and Technology, 47*(3), 373-387. doi:10.1080/0020739x.2015.1072881
Maddux, J. E., Sherer, M., & Rogers, R. W. (1982). Self-efficacy expectancy and outcome expectancy: Their relationship and their effects on behavioral intentions. *Cognitive Therapy and Research, 6*(2), 207-211. doi:10.1007/BF01183893
Mao, Y., White, T., Sadler, P. M., & Sonnert, G. (2016). The association of precollege use of calculators with student performance in college calculus. *Educational Studies in Mathematics, 94*(1), 69-83. doi:10.1007/s10649-016-9714-7
Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York: Routledge.
Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320-341. doi:10.1207/s15328007sem1103_2
Marton, F. (1981). Phenomenography: Describing conceptions of the world around us. *Instructional Science, 10*(2), 177-200. doi:10.1007/BF00132516

Marton, F., & Booth, S. A. (1997). *Learning and awareness*. Hillsdale, NJ: Lawrence Erlbaum.
Marton, F., & Säljö, R. (1976a). On qualitative differences in learning - II: Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology, 46*, 115-127. doi:10.1111/j.2044-8279.1976.tb02304.x
Marton, F., & Säljö, R. (1976b). On qualitative differences in learning I: Outcome and process. *British Journal of Educational Psychology, 46*, 4-11. doi:10.1111/j.2044-8279.1976.tb02980.x
Marton, F., & Säljö, R. (2005). Approaches to Learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning : Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 39-58). Edinburgh: The University of Edinburgh, Centre for Teaching, Learning and Assessment.
McDonald, R. P. (2011). *Test theory: a unified treatment*. New York: Routlege.
Michell, J. (1986). Measurement scales and statistics: a clash of paradigms. *Psychological Bulletin, 100*(3), 398–407. doi:10.1037/0033-2909.100.3.398
Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*, 355-383. doi:10.1111/j.2044-8295.1997.tb02641.x
Michell, J. (1999). *Measurement in psychology: a critical history of a methodological concept*. Cambridge: Cambridge University Press.
Michell, J., & Ernst, C. (1996). The Axioms of Quantity and the Theory of Measurement: Translated from Part I of Otto Hölder's German Text "Die Axiome der Quantität und die Lehre vom Mass". *Journal of Mathematical Psychology, 40*(3), 235-252. doi:10.1006/jmps.1996.0023
Mji, A. (2000). Conceptions of and approaches to learning mathematics. *Psychological Reports, 87*, 971-972. doi:10.2466/pr0.2000.87.3.971
Mji, A. (2003). A three-year perspective on conceptions of and orientations to learning mathematics of prospective teachers and first year university students. *International Journal of Mathematical Education in Science and Technology, 34*(5), 687-698. doi:10.1080/0020739031000148949
Moreno, V., & DiVesta, F. J. (1991). Cross-cultural comparisons of study habits. *Journal of Educational Psychology, 83*(2), 231-239. doi:10.1037/0022-0663.83.2.231
Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. New York: Chapman and Hall/CRC.
Mundia, L., & Metussin, H. (2019). Exploring factors that improve mathematics achievement in Brunei. *Studies in Educational Evaluation, 60*, 214-222. doi:10.1016/j.stueduc.2018.10.003
Nguyen, T. M. (2016). Learning approaches, demographic factors to predict academic outcomes. *International Journal of Educational Management, 30*(5), 653-667. doi:10.1108/ijem-06-2014-0085
Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.

Pajares, F., & Kranzler, J. H. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 20*(4). doi:10.1006/ceps.1995.1029

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *Journal of Educational Psychology, 86*(2), 193–203. doi:10.1037/0022-0663.86.2.193

Pajares, F., & Miller, M. J. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology, 42*(2), 190-198. doi:10.1037/0022-0167.42.2.190

Pearl, J., & Mackenzie, D. (2019). *The book of why: the new science of cause of effect*. United Kingdom: Penguin Random House.

Popper, K. (2002). *The Logic of Scientific Discovery* (6th ed.). London: Routledge.

Prediger, S., Bikner-Ahsbahs, A., & Arzarello, F. (2008). Networking strategies and methods for connecting theoretical approaches: first steps towards a conceptual framework. *ZDM Mathematics Education, 40*, 165–178. doi:10.1007/s11858-008-0086-z

Prudon, P. (2015). Confirmatory factor analysis as a tool in research using questionnaires: A critique. *Comprehensive Psychology, 4*. doi:10.2466/03.Cp.4.10

Rach, S., & Ufer, S. (2020). Which prior mathematical knowledge is necessary for study success in the university study entrance phase? Results on a new model of knowledge levels based on a reanalysis of data from existing Studies. *International Journal of Research in Undergraduate Mathematics Education, 6*(3), 375-403. doi:10.1007/s40753-020-00112-x

Raykov, T., & Marcoulides, G. A. (2016). Scale reliability evaluation under multiple assumption violations. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(2), 302-313. doi:10.1080/10705511.2014.938597

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*(11), 175–189. doi:10.1037/0022-0663.91.1.175

Rogosa, D. (1987). Casual models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics, 12*(2), 185-195. doi:10.3102/10769986012002185

Roick, J., & Ringeisen, T. (2018). Students' math performance in higher education: Examining the role of self-regulated learning and self-efficacy. *Learning and Individual Differences, 65*, 148-158. doi:10.1016/j.lindif.2018.05.018

Salmisto, A., Postareff, L., & Nokelainen, P. (2017). Relationships among civil engineering students’ approaches to learning, perceptions of the teaching–learning environment, and study success. *Journal of Professional Issues in Engineering Education and Practice, 143*(4). doi:10.1061/(asce)ei.1943-5541.0000343

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350-353. doi:10.1037/1040-3590.8.4.350

Schukajlow, S., Achmetli, K., & Rakoczy, K. (2019). Does constructing multiple solutions for real-world problems affect self-efficacy? *Educational Studies in Mathematics, 100*(1), 43-60. doi:10.1007/s10649-018-9847-y

Scriven, M. (1986). Evaluation as a paradigm for education research. In E. House (Ed.), *New Directions in Education Evaluation* (pp. 53–67). London: Falmer Press.

Shen, K.-M., Lee, M.-H., Tsai, C.-C., & Chang, C.-Y. (2016). Undergraduate students' earth science learning: Relationships among conceptions, approaches, and learning self-efficacy in Taiwan. *International Journal of Science Education, 38*(9), 1527-1547. doi:10.1080/09500693.2016.1198060

Sheu, H.-B., Lent, R. W., Miller, M. J., Penn, L. T., Cusick, M. E., & Truong, N. N. (2018). Sources of self-efficacy and outcome expectations in science, technology, engineering, and mathematics domains: A meta-analysis. *Journal of Vocational Behavior, 109*, 118-136. doi:10.1016/j.jvb.2018.10.003

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika, 74*(1), 107-120. doi:10.1007/s11336-008-9101-0

Socha, A., & Sigler, E. A. (2014). Exploring and "reconciling" the factor structure for the revised two-factor study process questionnaire. *Learning and Individual Differences, 31*, 43-50. doi:10.1016/j.lindif.2013.12.010

Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) Handout. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(6), 777-781. doi:10.1080/10705511.2016.1217487

Stes, A., De Maeyer, S., & Van Petegem, P. (2013). Examining the cross-cultural sensitivity of the revised two-factor study process questionnaire (R-SPQ-2F) and validation of a Dutch version. *PLoS One, 8*(1), e54099. doi:10.1371/journal.pone.0054099

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680. doi:10.1126/science.103.2684.677

Svensson, L. (2005). Skill in learning and organising knowledge. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 59-71). Edinburgh: The University of Edinburgh, Centre for Teaching, Learning and Assessment.

Tan, C. Y. (2018). Examining school leadership effects on student achievement: the role of contextual challenges and constraints. *Cambridge Journal of Education, 48*(1), 21-45. doi:10.1080/0305764x.2016.1221885

Tossavainen, T., Rensaa, R. J., & Johansson, M. (2019). Swedish first-year engineering students' views of mathematics, self-efficacy and motivation and their effect on task performance. *International Journal of Mathematical Education in Science and Technology,* 1-16. doi:10.1080/0020739x.2019.1656827

Trendler, G. (2009). Measurement theory, psychology and the revolution that cannot happen. *Theory & Psychology, 19*(5), 579-599. doi:10.1177/0959354309341926

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10. doi:10.1007/BF02291170

Tversky, A., Krantz, D. H., Suppes, P., & Luce, R. D. (1971). *Foundations of measurement volume I: Additive and polynomial representations*. New York: Academic Press.

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*(1), 89-101. doi:10.1016/j.cedpsych.2008.09.002

Wallace, B., & Kernozek, T. (2017). Self-efficacy theory applied to undergraduate biomechanics instruction. *Journal of Hospitality, Leisure, Sport & Tourism Education, 20*, 10-15. doi:10.1016/j.jhlste.2016.11.001

Wang, L., Li, X., & Li, N. (2014). Socio-economic status and mathematics achievement in China: A review. *Zdm, 46*(7), 1051-1060. doi:10.1007/s11858-014-0617-8

Whitelock-Wainwright, A., Gasevi, D., Wood, T., & Ryan, K. (preprint). A re-Evaluation of the Approaches to learning questionnaire. Retrieved from https://edarxiv.org/4degx/ (2020)

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of Achievement Motivation* (pp. 91-120). San Diego, California: Academic Press.

Williams, T., & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology, 102*(2), 453-466. doi:10.1037/a0017271

Yang, X., Leung, F., & Zhang, S. (2019). Junior secondary school students' conceptions of and approaches to learning mathematics and their relationships in mainland China. *Sustainability, 11*(9). doi:10.3390/su11092476

Yurt, E. (2014). The predictive power of self-efficacy sources for mathematics achievement. *Education and Science, 39*(176). doi:10.15390/eb.2014.3443

Zakariya, Y. F. (2019). Study approaches in higher education mathematics: Investigating the statistical behaviour of an instrument translated into Norwegian. *Education Sciences, 9*(3), 191. doi:10.3390/educsci9030191

Zakariya, Y. F. (2020). Investigating some construct validity threats to TALIS 2018 teacher job satisfaction scale: Implications for social science researchers and practitioners. *Social Sciences, 9*(4). doi:10.3390/socsci9040038

Zakariya, Y. F., Bjørkestøl, K., Nilsen, H. K., Goodchild, S., & Lorås, M. (2020). University students' learning approaches: an adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Education Evaluation*, 100816. doi:10.1016/j.stueduc.2019.100816

Zakariya, Y. F., Goodchild, S., Bjørkestøl, K., & Nilsen, H. K. (2019). Calculus self-efficacy inventory: Its development and relationship with approaches to learning. *Education Sciences, 9*(3), 170. doi:10.3390/educsci9030170

Zakariya, Y. F., Nilsen, H. K., Goodchild, S., & Bjørkestøl, K. (2020a). Assessing first-year engineering students' pre-university mathematics knowledge: Preliminary validity results based on an item response theory model. *Journal of Technology and Science Education, 10*(2), 259-270. doi:10.3926/jotse.1017

Zakariya, Y. F., Nilsen, H. K., Goodchild, S., & Bjørkestøl, K. (2020b). Self-efficacy and approaches to learning mathematics among engineering students: Empirical evidence for potential causal relations. *International Journal of Mathematical Education in Science and Technology*, 1-15. doi:10.1080/0020739X.2020.1783006

Zengin, Y. (2017). Investigating the use of the Khan academy and mathematics software with a flipped classroom approach in mathematics teaching. *Journal of Educational Technology & Society, 20*(2), 89-100.

Zientek, L. R., Fong, C. J., & Phelps, J. M. (2019). Sources of self-efficacy of community college students enrolled in developmental mathematics. *Journal of Further and Higher Education, 43*(2), 183-200. doi:10.1080/0309877x.2017.1357071

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ωH: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133. doi:10.1007/s11336-003-0974-7

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *J. Mod. Appl. Stat. Methods, 6*(1), 21-29. doi:10.22237/jmasm/1177992180

List of appendices

Appendix A: Revised Study Process Questionnaire (Norwegian version)

Dette spørreskjemaet inneholder noen spørsmål om holdningene du har til studiene dine og den måten du normalt studerer på. **Du bes her fokusere på emnet MA-178.** Det er ingen *riktig* måte å studere på. Det avhenger av hva som passer til din stil og av hvilke emner du studerer. Derfor er det viktig at du besvarer hvert spørsmål så ærlig som mulig.

Vennligst marker det passende alternativet med en (√), ved siden av det gitte spørsmålet på skjemaet kalt «Spørsmål om ‘generelle formål/svar-ark’». Bokstavene ved siden av hvert spørsmål står for følgende svar:

A - denne uttalelsen stemmer for min del aldri eller kun sjeldent

B – denne uttalelsen stemmer for min del noen ganger

C – denne uttalelsen stemmer for min del omtrent halvparten av tiden

D – denne uttalelsen stemmer for min del ofte

E – denne uttalelsen stemmer for min del alltid eller nesten alltid

	Spørsmål om ‘generelle formål/svar-ark’	A	B	C	D	E
1	Tidvis erfarer jeg at det å studere gir meg en dyp, personlig tilfredsstillelse					
2	Jeg erfarer at jeg må arbeide tilstrekkelig med et tema, slik at jeg kan trekke mine egne konklusjoner før jeg blir fornøyd					
3	Målet mitt er å bestå emnet og samtidig arbeide så lite som mulig					
4	Jeg studerer kun seriøst det som meddeles i undervisningen eller i emnebeskrivelsene					
5	Det er min erfaring at nesten alle tema kan være meget interessante, straks jeg setter meg inn i dem					
6	Jeg synes de fleste nye tema er interessante og bruker ofte ekstra tid på å forsøke og skaffe til veie mer informasjon om dem					

7	Jeg synes ikke at emnet jeg tar er veldig interessant så jeg arbeider minimalt					
8	Jeg synes at det å studere akademiske fag til tider kan være like spennende som en god bok eller en god film					
9	Jeg tester meg selv i viktige tema inntil jeg forstår dem fullt ut					
10	Jeg erfarer at jeg kan klare meg gjennom de fleste vurderingsformer ved å memorere viktige avsnitt, fremfor å gjøre forsøk på å forstå dem					
11	Generelt begrenser jeg mine studier til det som spesifikt er oppgitt, ettersom jeg synes det er unødvendig å gjøre noe ekstra					
12	Jeg arbeider hardt med mine studier fordi jeg finner stoffet interessant					
13	Jeg bruker mye av min fritid på å finne ut mer om interessante tema som har blitt diskutert i ulike undervisningssituasjoner					
14	Jeg synes ikke det er til hjelp å studere emner i dybden. Det forvirrer meg og tiden kastes bort, når alt man behøver er grunnleggende kjennskap til de ulike temaene					
15	Jeg mener at forelesere ikke bør forvente at studentene bruker betydelig tid på å studere stoff som alle vet at det ikke vil bli eksaminert i					
16	Jeg møter til undervisning med spørsmål i tankene, og som jeg ønsker at blir besvart					
17	Jeg gjør et poeng ut av å se på mesteparten av den foreslåtte litteraturen som knyttes til forelesningene					
18	Jeg ser ikke noe poeng i læringsmateriale som det mest sannsynlig ikke vil bli eksaminert i					
19	Jeg erfarer at den beste måten å bestå eksamen på, er å prøve og huske svar på tilsvarende spørsmål					

Appendix B: Revised study process questionnaire (English translation of the Norwegian version)

This questionnaire has a number of questions about your attitudes towards mathematics and your usual way of studying it. **Please focus here on the course MA-178**. There is no *right* way of studying. It depends on what suits your own style and the course you are studying. It is accordingly important that you answer each question as honestly as you can.

Place the mark (√) at the appropriate option to each statement. The letters alongside each number stand for the following response.

A—this item is *never* or *only rarely* true of me

B—this item is *sometimes* true of me

C—this item is true of me about *half the time*

D—this item is *frequently* true of me

E—this item is *always* or *almost always* true of me

	Statement on approaches to learning mathematics	A	B	C	D	E
1	I find that at times studying gives me a feeling of deep personal satisfaction.					
2	I find that I have to do enough work on a topic so that I can form my own conclusions before I am satisfied.					
3	My aim is to pass the course while doing as little work as possible.					
4	I only study seriously what's given out in class or in the course outlines.					
5	I feel that virtually any topic can be highly interesting once I get into it.					
6	I find most new topics interesting and often spend extra time trying to obtain more information about them					
7	I do not find my course very interesting so I keep my work to the minimum.					
8	I find that studying academic topics can at times be as exciting as a good novel or movie.					

9	I test myself on important topics until I understand them completely.					
10	I find I can get by in most assessments by memorising key sections rather than trying to understand them.					
11	I generally restrict my study to what is specifically set as I think it is unnecessary to do anything extra.					
12	I work hard at my studies because I find the material interesting.					
13	I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes.					
14	I find it is not helpful to study topics in depth. It confuses and wastes time, when all you need is a passing acquaintance with topics.					
15	I believe that lecturers shouldn’t expect students to spend significant amounts of time studying material everyone knows won’t be examined.					
16	I come to most classes with questions in mind that I want answering					
17	I make a point of looking at most of the suggested readings that go with the lectures.					
18	I see no point in learning material which is not likely to be in the examination.					
19	I find the best way to pass examinations is to try to remember answers to likely questions.					

Appendix C: Calculus self-efficacy inventory (Norwegian)

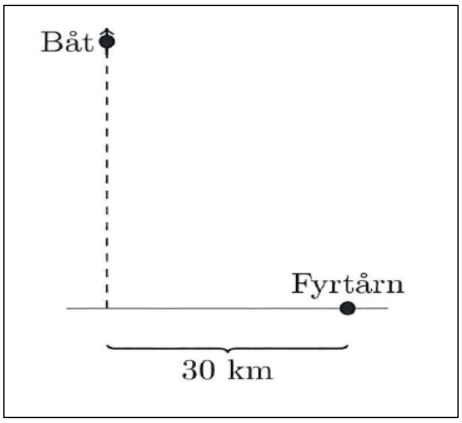
Vennligst svar på spørsmålene ærligst mulig. Svarene behandles konfidensielt og vil kun bli brukt i forskningsøyemed.

I. Kjønn: ☐ Mann ☐ Kvinne [Marker med et kryss (X)]

Punktene under utgjør en liste over ulike oppgaver fra Kalkulus-pensumet. I kolonnen "Tiltro" bes du om å angi hvor stor tiltro du har til at du hadde maktet å løse hver av oppgavene **her og nå**. Angi graden av tiltro med et tall fra 0 til 100, ut ifra skalaen under:

0	10	20	30	40	50	60	70	80	90	100
Kan overhodet ikke løse den				Moderat sikker på å løse den					Meget sikker på å løse den	

Oppgave	Hvor stor tiltro har du til at du kan løse hver av oppgavene under, **her og nå**?	Tiltro (0 – 100)
1	Regn ut $(1-i)^{666}$	
2	Regn ut grenseverdien $$\lim_{x\to 1}\frac{1-\cos(1-x^2)}{x^2-2x+1}$$	
3	Gitt funksjonen $f(x)=\ln(2x^2-3x+2)$. Finn definisjonsmengden til funksjone f?	
4	Gitt funksjonen $f(x)=x+e^x$. Verdien av c som tilfredsstiller konklusjonen av **Middelverditeoremet, Mean value Theorem** på intervallet $[0,1]$ er hva?	
5	En funksjon er gitt ved $f(x)=e^{-x}(x^2+4x+1)$. Finn ved regning alle ekstremalpunktene til f og bestem om de er toppunkt eller bunnpunkt.	

6	En kurve er gitt ved $x = y^2 - x^2y - 1$. Bruk implisitt derivasjon til å finne et uttrykk for y'.	
7	En partikkel beveger seg på grafen til $f(x) = \ln x,\ x > 1$ med en hastighet 2 enheter per sekund, målt på x −aksen. Hvor raskt endrer avstanden fra origo til partikkelen seg i det partikkelen passerer der $x = e$?	
8	En båt kjører ut fra kysten i retning rett mot nord. Det står et fyrtårn 30 km øst for punktet på kysten hvor båten la ut ifra. På et tidspunkt observeres det med radar fra fyrtårnet at båten er nøyaktig 50 km fra fyrtårnet og at avstanden mellom båten og fyrtårnet øker med 3 meter per sekund. Hvor fort kjører båten på dette tidspunktet? Gi svaret i km per time. 	
9	Regn ut integralene $$\int \frac{x-7}{x^2+x-6}\,dx$$	
10	Regn ut integralene	

	$$\int_0^{\infty} x^2 e^{-x} dx$$	
11	Finn $$\int \frac{1}{4x^2+1} dx$$	
12	En flate er avgrenset av funksjonen $f(x) = \frac{1}{3} e^{x^2}$ der $0 \leq x \leq 2$ og x –aksen. Et kar lages ved å rotere flaten om x-aksen. Finn volumet av karet.	
13	Vi skal la $f(x) = \sqrt{x} \cdot e^{\sqrt{x}}$ og $D_f = (0,4]$ i denne oppgava. Funksjonen h er definert på [0, 2] ved at $h(x) = \int_1^{x^2} f(t)dt$. Bestem $h'(x)$.	

Appendix D: Calculus self-efficacy inventory (English)

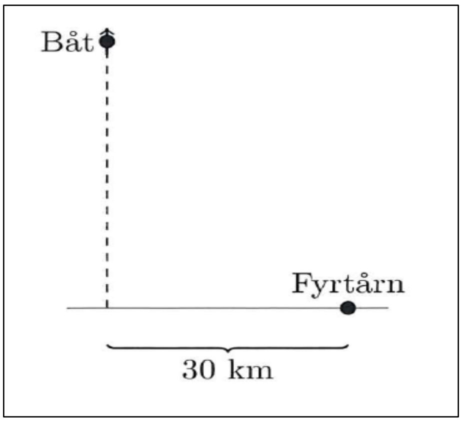
Please answer the questions below as honestly as possible. Your answers will be treated with confidentiality and will only be used for research purposes.

I. Sex: ☐ Male ☐ Female [mark with a cross (X)]

Below you will find a set of different tasks from the Calculus syllabus. In the column "Confidence" you are asked to state how much confidence you have that you could manage to solve each task – **here and now**. Please provide your level of confidence with a number from 0 to 100, using the scale below:

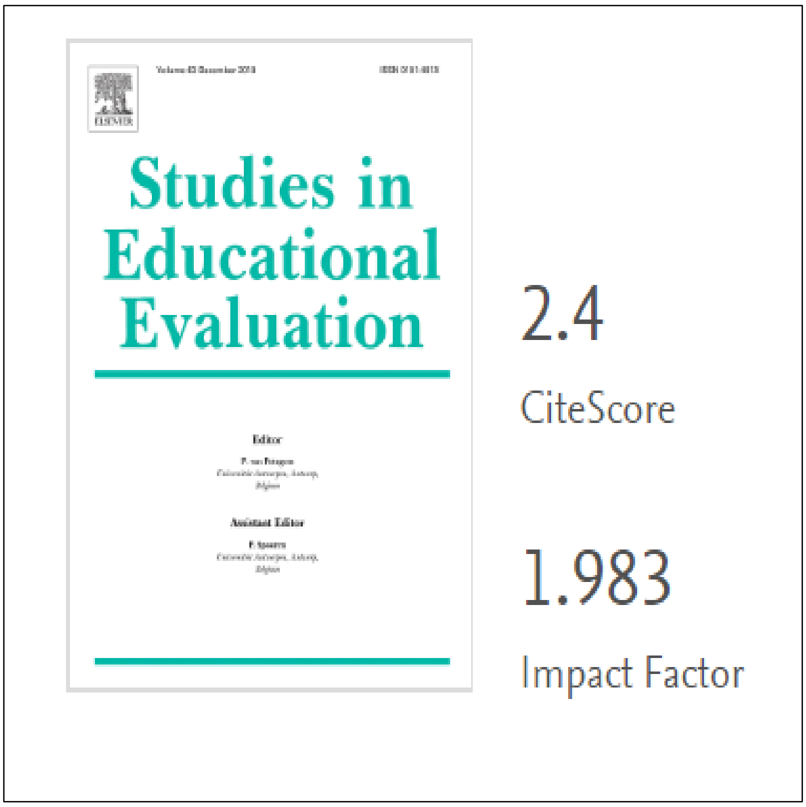
0	10	20	30	40	50	60	70	80	90	100
No possibility that I could solve this				Moderately confident that I could solve this					Totally confident that I could solve this	

Task	How confident are you that you can solve each of these problems **right now**?	Confidence (0 – 100)
1	Calculate $(1-i)^{666}$	
2	Calculate the value of the limit $$\lim_{x\to 1}\frac{1-\cos(1-x^2)}{x^2-2x+1}$$	
3	Given the function $f(x)=\ln(2x^2-3x+2)$. What is the domain of the function f?	
4	Given the function $f(x)=x+e^x$. What is the value of c that satisfies the result of the **Mean Value Theorem** in the interval $[0,1]$?	
5	A function is given so that $f(x)=e^{-x}(x^2+4x+1)$. Find, by calculating, all the turning points of f and determine the maximum and minimum points.	
6	A curve is given by the equation $x=y^2-x^2y-1$. Use implicit derivation to find an expression for y'.	

7	A particle moves on the graph of $f(x) = \ln x,\ \ x > 1$ with a speed 2 units per second, measure along the x-axis. How fast does the distance from the origin to the particle change when the particle is $x = e$?	
8	A boat sails from the coast towards the North. A lighthouse stands 30 km east from the point on the coast where the boat departed. At one moment in time, radar from the lighthouse shows that the boat is exactly 50 km from the lighthouse and the distance between the boat and the lighthouse is increasing at 3 metres per second. How fast is the boat moving at that moment? Give the answer in km per hour. 	
9	Solve the integral $$\int \frac{x-7}{x^2+x-6}dx$$	
10	Calculate the integral $$\int_0^{\infty} x^2 e^{-x} dx$$	
11	Find $$\int \frac{1}{4x^2+1}dx$$	

12	A surface is bounded by the function $f(x) = \frac{1}{3}e^{x^2}$ where $0 \leq\ x \leq\ 2$ and the x-axis. A vessel is made by rotating the surface around the x-axis. Find the volume of the vessel.	
13	In this task, let $f(x) = \sqrt{x} \cdot e^{\sqrt{x}}$ and $D_f\ = (0,4]$. The function h is defined on $[0, 2]$, with $h(x) = \int_1^{x^2}\ f(t)dt$. What is $h'(x)$?	

Paper I



Zakariya, Y. F., Bjørkestøl, K., Nilsen, H. K., Goodchild, S., & Lorås, M. (2020). University students' learning approaches: an adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Education Evaluation*, 100816. doi:10.1016/j.stueduc.2019.100816

Studies in Educational Evaluation 64 (2020) 100816

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

University students' learning approaches: An adaptation of the revised two-factor study process questionnaire to Norwegian

Yusuf F. Zakariya[a,*], K. Bjørkestøl[a], H.K. Nilsen[a], S. Goodchild[a], M. Lorås[b]

[a] Department of Mathematical Science, University of Agder, Kristiansand, Norway
[b] Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

Keywords:
Confirmatory factor analysis
R-SPQ-2F
Validity
Reliability
Learning approaches

ABSTRACT

This paper reports a Norwegian validation study of a widely used instrument to measure students' approaches to learning, namely, Bigg's revised two-factor study process questionnaire (R-SPQ-2F). Its cultural sensitivity and psychometry evaluations have provoked rigorous discussion among educators in different languages. A survey design was adopted involving 253 undergraduate engineering students across two universities. Confirmatory factor analyses were used to test six models hypothesized to reflect the factor structures of R-SPQ-2F and unidimensionality of its subscales. The results showed appropriate fits of a two-factor first-order model with 10 items measuring deep approach and 9 items measuring surface approach subscales. The reliability was found to be high with coefficients of .81, .72 and .63 on deep subscale, surface subscale and the whole instrument respectively. Findings may be interpreted as evidence of cultural sensitivity of the instrument and more validation studies were recommended.

1. Introduction

The increase in number and diversity of higher education students coupled with huge investment on the parts of government, parents, educational stakeholders and students have prompted enormous research into undergraduate students' learning experience. An important aspect of students learning that has attracted attention of education researchers over the last decades is their learning approaches (e.g., Fryer & Vermunt, 2018; Maciejewski & Merchant, 2016). Approaches to learning in higher education (HE) connotes predispositions adopted by an individual when presented with learning materials and strategies used to process the learning contents (Baeten, Kyndt, Struyven, & Dochy, 2010). A long-standing categorization of learning approaches into notions of "deep" and "surface" was introduced by Ference Marton and colleagues over 40 years ago.

Marton and Säljö developed the students' approaches to learning (SAL) theory from their qualitative clinical experimental series of studies (Marton & Säljö, 1976a, 1976b) on Swedish undergraduate students' approaches to reading, understanding and answering questions based on some presented passages of prose and newspaper articles. The experiments were aimed at exploring qualitative differences in the presented materials and describing practical differences in learning processes. In these experiments, they utilized the term "approaches to learning" to connote the *processes* adopted by the students, prior to the experiments which directly influence their learning outcome. The series of experiments resulted in a categorization of students' learning processes into deep and surface approaches.

A deep approach learner processes information with the intent of discovering the meaning of intended content of the material while a surface approach learner is preoccupied with the discourse or the text itself with little or no attention to the intended meanings. More recently, Biggs (2012) while describing surface and deep approaches to learning posited that the surface approach to learning "refers to activities of an inappropriately low cognitive level, which yields fragmented outcomes that do not convey the meaning of the encounter" and the deep approach to learning "refers to activities that are appropriate to handling the task so that an appropriate outcome is achieved." (p.42).

Measurement of students' approaches to learning is an aspect of instruction in HE that has attracted attention for the past 45 years. Questions like what should be measured in SAL?, how should it be measured?, and how many subcategories should SAL measuring instrument contain?, etc., have been investigated extensively (e.g., Kember, 1990). John Bigg's revised two-factor study process questionnaire R-SPQ-2F has been identified among the most widely studied instruments for measuring approaches to learning (e.g., Lake, Boyd, & Boyd, 2017; López-Aguado & Gutiérrez-Provecho, 2018). Similar instruments are the approaches and study skills inventory for students (ASSIST) and revised approaches to studying inventory (RASI) that

* Corresponding author.
E-mail address: yusuf.zakariya@uia.no (Y.F. Zakariya).

https://doi.org/10.1016/j.stueduc.2019.100816
Received 22 May 2019; Received in revised form 6 September 2019; Accepted 7 September 2019
Available online 09 October 2019

were developed and validated in different languages (e.g., Diseth, 2001; Tait, Entwistle, & McCune, 1998; Valadas, Gonçalves, & Faísca, 2010). Meanwhile, R-SPQ-2F has advantage over ASSIST with regards to its concise length and it is more readily interpretable than the RASI because of its lower number of primary latent factors.

However, cultural sensitivity of R-SPQ-2F when adapted to different languages has generated heated debates among researchers (López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014). In most instances, only two latent factors as opposed to four hypothesized by Biggs, Kember, and Leung (2001) have been reported to be the best explanation for the factor structure of the instrument (e.g., López-Aguado & Gutiérrez-Provecho, 2018). Contrary to Biggs et al. (2001), a handful of studies also recommended deletion of some items from the original instrument in order to achieve model fits (e.g., Socha & Sigler, 2014). These contrasting findings have created knowledge gaps for more studies on the cultural sensitivity of the instrument. It is therefore necessary to validate the Norwegian version of R-SPQ-2F before applying it to our university students. The main purpose of this study is to confirm the underlying factor structure of R-SPQ-2F and establish its reliability estimates using appropriate psychometric analysis.

2. Literature review

2.1. Factor structures of R-SPQ-2F

Psychometric properties such as validity and reliability of R-SPQ-2F have been studied extensively and the results well documented (Biggs et al., 2001; Chan & Sheung Chan, 2010; Weller et al., 2013). In Biggs et al. (2001), validity, reliability and dimensionalities of R-SPQ-2F were investigated involving 495 university students across various departments in a university in Hong Kong. The unidimensionality of each substructure – deep motive (DM), deep strategy (DS), surface motive (SM) and surface strategy (SS) – was investigated by conducting confirmatory factor analysis (CFA) which established the homogeneity of each 5-item subscale. Two models were hypothesized and tested using CFA to explain the factor structures of R-SPQ-2F. The first model (see, Fig. 1A) was a first-order four-factor model – DM, DS, SM and SS – with partial covariance and five indicators on each latent variable. The results showed a good fit with standardized root mean square residual (SRMR) = .058, comparative fit index (CFI) = .904 and correlations of .93, .70 and -.18 between DM and DS, SS and SM, and DM and SM respectively. The second model (see, Fig. 1B) was as well a first-order two-factor model – deep and surface – with two indicators each DM and DS, SM and SS respectively got by summing items corresponding to the subscales. The results also showed a good fit with SRMR = .015, CFI = .992 and correlation −.23 between deep and surface factors.

A reliability check was conducted and Cronbach's alpha coefficients of .62, .63, .72 and .57 were reported for DM, DS, SM and SS respectively. Further, acceptable Cronbach's alpha coefficients of .73 and .64 were also reported for the 10-item deep approach (DA) and Surface Approach (SA) factors respectively (Biggs et al., 2001). In a similar corroborative empirical study involving 404 students of higher diplomas and associate degrees in Hong Kong, Chan and Sheung Chan (2010) reported much higher Cronbach's alpha coefficients of .70, .74, .70, .65, .85 and .80 for DM, DS, SM, SS, DA, and SA respectively. More so, Weller et al. (2013) conducted an exploratory factor analysis (EFA) using maximum likelihood (ML) coupled with CFA after some changes in the wordings of R-SPQ-2F to suit their research field. The results made a perfect match of the two-factor extracted as in the original instrument with a considerable internal consistency and Cronbach's alpha values of .74 and .83 for DA and SA respectively.

2.2. Cultural sensitivity of R-SPQ-2F

The cultural sensitivity of R-SPQ-2F has stirred up debates among educationists in recent time especially when adapted into Spanish (Justicia, Pichardo, Cano, Berbén, & De la Fuente, 2008), Turkish (Önder & Besoluk, 2010), Japanese (Fryer, Ginns, Walker, & Nakao,

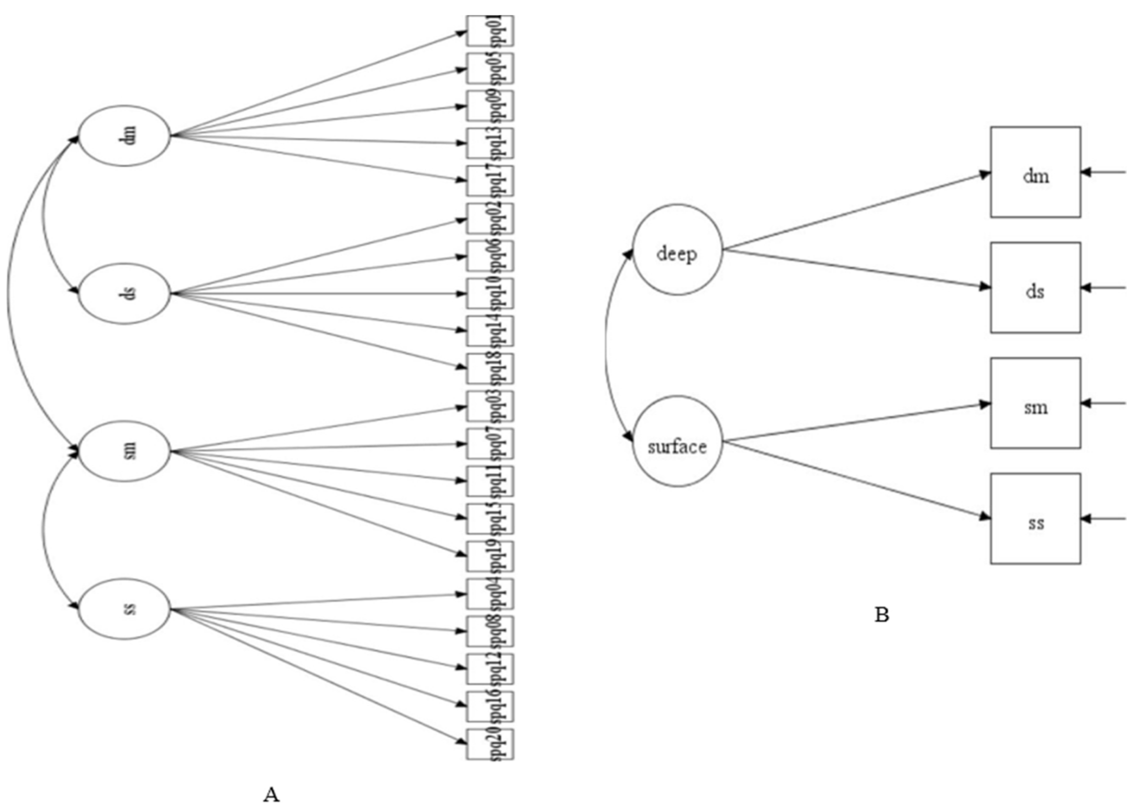


Fig. 1. Models 1 and 2 as hypothesized by Biggs et al. (2001).

2012), Dutch (Stes, De Maeyer, & Van Petegem, 2013), Chinese (Xie, 2014), and Arabic (Shaik et al., 2017). The results of these studies have ended in different conclusions with most studies proposing a two-factor R-SPQ-2F without any further subcategories into motive and strategy. In an attempt to investigate this phenomena, Leung, Ginns, and Kember (2008) conducted an empirical study on two independent samples of 1146 university students in Australia and 1266 students in Hong Kong. Their results showed no significant difference in the description of students approaches to learning in both countries, the range of developing approaches from surface to deep is common and the Cronbach's alpha coefficients ranged from 0.64 (SS) to 0.74 (SM) and 0.70 (SS) to 0.77 (DM) for both Hong Kong and Australian samples respectively. The hypothesized two-factor structural model was confirmed using CFA which gave a good fit for both samples (Leung et al., 2008). This is one of the few studies that have confirmed no cultural sensitivity of R-SPQ-2F across different cultural settings.

On the other hand, Justicia et al. (2008) were among the earlier researchers to provoke the discussion on cultural sensitivity of R-SPQ-2F. In their empirical study, data were collected from two independent samples of 314 and 522 university students. The first sample composed of mainly year one education students (used for EFA) and second sample composed of 274 and 248 final year students of education and psychology respectively (used for CFA). The R-SPQ-2F was translated to Spanish employing back-translation coupled with some modifications to cater for cultural differences. Their analysis was rigorous including EFA (both PCA and PFA – Principal Factor Analysis), CFA, item polychoric correlations to cater for multivariate normality and comparing other models. The final results confirmed two-factor structures for R-SPQ-2F, and no empirical evidence was found for differentiating between motive and strategy subscales. A corroborative result for best fit of two underlying factor structures for R-SPQ-2F was as well reported in the Turkish version of the instrument (Önder & Besoluk, 2010). Evidence of reliability was also provided with Cronbach's alpha coefficients of .78 and .74 for deep and surface dimensions respectively. The Japanese (Fryer et al., 2012), Dutch (Stes et al., 2013) and Arabic (Shaik et al., 2017) versions also reported two underlying factor structures for the R-SPQ-2F in their respective studies with little conceptual variations in deep and surface approaches.

A study that stood out almost completely was the report of Immekus and Imbrie (2010) involving two cohorts (A = 1490 and B = 1533) of university students in the United States of America. The reliability estimates were .81 and .80 (cohort A) and .81 and .78 (cohort B) for deep and surface approach subscales respectively. The interesting part was the factor analysis results. There was no empirical evidence for neither the two-factor nor for the four-factor structures of the R-SPQ-2F in the cohort A. However, a four-factor model was found fit after deleting 5 items. This was later confirmed using CFA on cohort B and found to have a good fit with acceptable statistics and the final four-factor items considerable overlapped with Bigg's et al. 2001 initial substructures (Immekus & Imbrie, 2010). In an attempt to reconcile between these variant reports on latent structures of R-SPQ-2F, Socha and Sigler (2014) conducted an empirical study involving 868 university students and compared 8 statistical models. Rather than solving the problem, they also came up with a two-factor best description of R-SPQ-2F at a cost of deleting two items (Socha & Sigler, 2014).

3. Methods

3.1. Participants

A total of 253 year-one university engineering and computer science students participated in this study. This comprised 168 males and 72 females distributed across two universities in Norway and age range of 20–23 years. 13 students did not indicate their gender. An effective sample of size of 253 was realized after subtracting ten missing cases in the main data. Despite the sample size was smaller than envisaged due to general attitudes of undergraduate students towards responding to questionnaires, it conforms with the recommendations of Monte Carlo simulation studies reported in (Gagne & Hancock, 2006; Wolf, Harrington, Clark, & Miller, 2013). This was based on the many elements such number of factors ($\leq$ 4), expected factor loading ($\leq$ .8), number of indicators per factor ($\leq$ 10), expected power ($\geq$ .8), expected ratio of χ^2 -value to df ($\leq$ 4), etc.

3.2. Materials

R-SPQ-2F was translated independently by two Norwegian first language associate professors of mathematics education. Comparison of translated versions was done, and agreements were reached on the appropriate choices of words. A back-translation to English was conducted by an English professor of mathematics education who has spent about 15 years in Norway. The back translation was compared with the original English version and minor corrections were made to cater for cultural language differences. The instrument was then converted to electronic form using SurveyXact and paper version was printed for back-up.

3.3. Procedure for data collection

Electronic version of consent forms was sent to the students via their university emails followed by a class visit for a presentation on the project. In the presentation, we gave a brief description of our project to the students and stressed the importance of their involvement in the research. At this instance, some students filled-out the paper version of the consent forms. A week after, we paid another visit with paper version of the translated R-SPQ-2F, gave a 5-minute presentation on the questionnaires and some students as well completed the paper version. This was preceded by distribution of R-SPQ-2F electronic version via emails. We gave a time frame of about three weeks to receive responses accompanied with occasional reminders. The response rate was about 35% of the total population. The low response rate could be ascribed to the general attitudes of undergraduate students towards completing questionnaires as well the busy schedules of most of the students at the time.

3.4. Procedure for data analysis

The collected data from both paper and electronic versions of R-SPQ-2F were merged, screened, relabeled, coded and saved in ASCII format. Confirmatory factor analysis was used to test six models and the results were reported in the current article. The first CFA was used to confirm model 1 proposed by Biggs et al. (2001) using weighted least square mean and variance adjusted (WLSMV) estimator in Mplus version 8.3 (Muthén & Muthén, 1998-2017; Muthén and Muthén, 1998). WLSMV was utilized as it is robust enough to perform well on analysis of ordinal data (in which basic assumptions of normality, absence of kurtosis and skewness are violated), presence of missing data and small sample size as compared to ML and others (Brown, 2015; Suh, 2015). The second CFA was used to test model 2 proposed by Biggs et al. (2001). The default ML estimator was used for this model because summing the indicators scores has inflated the categories which make it too cumbersome of WLSMV to handle. Model 3 was a modification of model 2 containing four first-order factors – DM, DS, SM and SS – measured by five indicators each and two second-order factors – deep and surface – hierarchical model.

Model 4 was a proposed modified version of model 3 containing two first-order factors – deep and surface – model measured by ten and nine indicators respectively. Models 5 and 6 were single-factor models used to check the unidimensionality of items in deep and surface subscales. WLSMV estimator was used in the analysis of models 3-6. Cronbach alpha coefficient estimate for the reliability of the instrument was not used because it depends on Pearson correlations which requires normality assumption for accurate estimates. Rather, the internal consistency of R-SPQ-2F and its subscales was checked using Raykov and Marcoulides' formula which have been confirmed to performed more efficiently than Cronbach alpha under violations of multiple assumptions (Raykov & Marcoulides, 2016). For instructional purposes, the data used for this study are available upon request and Mplus syntax codes as well the English final version of R-SPQ-2F are enclosed in the appendices. The Norwegian version of R-SPQ-2F is available upon request from the corresponding author.

3.5. Criteria for assessing a model fit

Apart from a non-significant χ^2 -value, there are a number goodness of fits (GOF) indices proposed to assess the optimality of approximate prediction of sample matrix by a CFA model. Popularly reported indices in educational studies are: TLI-Tucker-Lewis index (Tucker & Lewis, 1973), RMSEA-root mean square error of approximation (Steiger and Lind, 1980 in Steiger, 2016), SRMR (Jöreskog & Sörbom, 1988), and CFI (Bentler, 1990). For both CFI and TLI a value 1.00 indicates a perfect model fit while values close to or greater than 0.90 indicate a good fit (Bentler, 1990; Hu & Bentler, 1999). A cut-off RMSEA value of less than or equal to 0.06 was proposed by Hu and Bentler (1999) for a good model fit. Other experts (e.g., Browne & Cudeck, 1992) have proposed RMSEA values between 0.00 to 0.05 and 0.05 to 0.08 as depicting a good and an adequate model fits respectively. A model with RMSEA value between 0.08 to 0.10 was characterized as having a "mediocre fit" while models with value greater than 0.10 should be rejected (MacCallum, Browne, & Sugawara, 1996). In the case of SRMR, a value less than or equal to .08 was suggested by Hu and Bentler (1999) as an indicator of a good fit.

In practice, methodologists and researchers do not take the cut-off values of GOF indices as a rule of thumb. In fact, a close look at the work of Hu and Bentler (1999) revealed that their cut-off criteria are not generalizable especially when other estimators e.g. WLSMV apart from ML are used and more than five indicators per factors are involved in the instrument (Marsh, Hau, & Wen, 2004). Further, Hu and Bentler (1998), 1999) criteria have been considered unrealistic for most social sciences research especially when the data involved ordinal scales with multiple violations of assumptions (Marsh et al., 2004). It is therefore helpful, and of course the criteria adopted in the current study, to utilize a combination of the indices with some relaxation in cut-off values coupled with significant level of indicator factor loadings and interpretability of other parameter estimates.

4. Results

4.1. Analysis of hypothesized model 1 (Biggs et al., 2001)

The first-order four-factor model of Biggs et al. (2001) was subjected to CFA and the results are presented in Tables 1 and 2.

The results in Table 1 show a poor fit of model 1. This is evident with a significantly high χ^2 -value (167, N = 253) = 609.79, $p < .05$, and none of the fit indices is within the recommended acceptable range. In fact, the problem of this model is worse than the out of range fit indices. The latent variable covariance matrix is not positive definite (see, Table 2) which renders the model non-admissible. This was because of the presence of Heywood cases in form of standardized correlations great than one (1.018 and 1.048) between latent variables DM and DS, SM and SS respectively. This is interpreted to be an evidence of over-factoring in the model and gross misspecifications that are suggestive of redundant latent factors with high multicollinearity (Brown, 2015; Byrne, 2012). One may argue that the nonpositive definite matrix was due to pairwise estimations (one by one correlation) involved in the computation of polychoric correlation matrix used in the analysis of ordinal data. In order to clear this doubt, the estimator was changed to ML and the analysis was run again. The result is no different from the one reported in Tables 1 and 2. Hence, it can be deduced that the hypothesized model 1 is not descriptive enough of the data and therefore rejected.

Table 1
Mplus output of model 1: Selected GOF statistics.

Tests of model fits	
Chi-Square Test of Model Fit	
Value	609.786
Degrees of freedom	167
p-value	0.0000
RMSEA (Root Mean Square Error of Approximation)	
Estimate	0.102
90 Percent C.I.	0.094 0.111
Probability RMSEA < = .05	0.000
CFI/TLI	
CFI	0.710
TLI	0.670
Number of Free Parameters	103
SRMR (Standardized Root Mean Square Residual)	
Value	0.100

Table 2
Estimated correlation matrix for the latent variables.

	DM	DS	SM	SS
DM	1.000			
DS	1.018	1.000		
SM	−0.602	0.000	1.000	
SS	0.000	0.000	1.048	1.000

This finding, though contrary to the hypothesized model proposed by Biggs et al. (2001) and those who confirmed it (e.g., Merino & Kumar, 2013; Xie, 2014) it does conform to the results of non-admissible solutions reported in many studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014; Stes et al., 2013). Moreover, some of the studies that confirmed admissible solutions for the four-model (e.g., Xie, 2014) also found high correlation coefficients between DM and DS, SM and SS which are suggestive of an over-factored model. They therefore concluded their studies with a two-factor explanation of the instrument (e.g., Merino & Kumar, 2013; Xie, 2014).

4.2. Analysis of hypothesized model 2 (Biggs et al., 2001)

The first-order two-factor model of Biggs et al. (2001) was subjected to CFA and the results are presented in Table 3 and Fig. 2.

The results in Table 3 appear to show a good model fit from the perspective of GOF indices. The χ^2 -value (1, N = 253) = 3.269, $p > .05$ was not significant and all the fit indices are within recommended acceptable range except RMSEA. However, the main

Table 3
Mplus output of model 2: Selected GOF statistics.

Tests of model fits	
Chi-Square Test of Model Fit	
Value	3.269
Degrees of Freedom	1
p-value	0.0706
CFI/TLI	
CFI	0.991
TLI	0.944
Number of Free Parameters	13
RMSEA (Root Mean Square Error of Approximation)	
Estimate	0.095
90 Percent C.I.	0.000 0.217
Probability RMSEA < = .05	0.160
SRMR (Standardized Root Mean Square Residual)	
Value	0.017

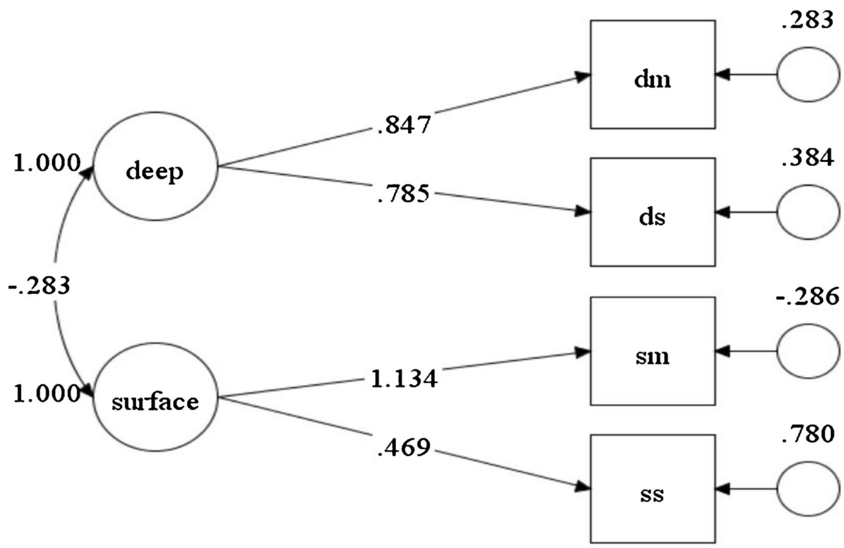


Fig. 2. Model 2 diagram with standard estimated parameters.

problem of this model is that the residual covariance matrix was not positive definite which renders the model non-admissible. This was because of the presence of Heywood case in form of negative unique variance of SM, see Fig. 2. This is suggestive of model gross misspecifications as positive definite variance/covariance matrix is a necessary condition for an admissible model (Brown, 2015; Kolenikov & Bollen, 2012). Hence, it can be deduced that the hypothesized model 2 as well is not descriptive enough of our data and therefore rejected.

This finding as well, though contrary to hypothesized model proposed by Biggs et al. (2001) it does conform to the results of non-admissible solutions reported in many studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014; Stes et al., 2013). The negative unique variance found in the SM indicator completely overlapped with the finding of Socha and Sigler (2014) who also found negative error disturbance in both SM and SS indicators.

4.3. Analysis of proposed hierarchical model 3

Observed methodological issues in terms of adding up scores on component items to make indicators in model 2 coupled with its within range accompanied GOF indices prompted the test of a hierarchical four-factor model. It consists two first-order and two second-order factors tested at item levels contrary to aggregating scores used by Biggs et al. (2001). This proposed hierarchical model also relied on previous studies which have tested similar models and found admissible solutions (e.g., Justicia et al., 2008). The results are presented in Table 4 and Fig. 3.

Table 4
Mplus output of model 3: Selected GOF statistics.

Tests of model fits	
Chi-Square Test of Model Fit	
Value	521.114
Degrees of Freedom	168
p-value	0.0000
CFI/TLI	
CFI	0.769
TLI	0.739
Number of Free Parameters	102
RMSEA (Root Mean Square Error of Approximation)	
Estimate	0.091
90 Percent C.I.	0.082 0.100
Probability RMSEA < = .05	0.000
SRMR (Standardized Root Mean Square Residual)	
Value	0.081

The results in Table 4 show a poor fit of the proposed hierarchical model. This is evident with a significantly high χ^2 -value (168, N = 253) = 521.11, $p < .05$, and none of the fit indices is within the recommended acceptable range. Here again, the latent variable covariance matrix is not positive definite which renders the model non-admissible. This was because of presence of Heywood cases in form of negative unique variance in latent variables DM and DS, SM and SS. This is interpreted to be an evidence of over-factoring in the model and gross misspecifications that are suggestive of redundant latent factors (Brown, 2015; Byrne, 2012). Therefore, it can be deduced that the hypothesized model 3 is not descriptive enough of the data and rejected. This finding also corroborated previous studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Merino & Kumar, 2013; Socha & Sigler, 2014) who have also reported poor fit as well as non-admissible solutions of this model.

4.4. Analysis of proposed model 4

The over-factoring observed in model 3 was corrected by collapsing latent variables DM with DS and SM with SS to form a hypothesized two-factor model and tested at item level. The results were presented in Table 5 and Fig. 4.

The results in Table 5 (with item 8) show a poor fit of the proposed two-factor model. This is evident with a significantly high χ^2 -value (169, N = 253) = 522.18, $p < .05$, and none of the fit indices is within the recommended acceptable range. In fact, all the GOF indices are the same with ones obtained in model 3 except a slight change in χ^2 -value. However, the model solution was admissible with a positive definite variance/covariance matrix. We investigated the estimated standardized factor loadings and found that item 8 has an extremely small nonsignificant loading (.06, $p > .05$) on surface approach. This item was removed as its contribution is negligible to the instrument.

The analysis was repeated and the obtained are results in Table 5 (without item 8). This showed an admissible solution with reduced χ^2 -value (151, N = 253) = 377.68, significant $p < .05$ with $\chi^2/df < 3$. All factor loadings are significant (see, Fig. 4), SRMR ($\leq$ .08), CFI/TLI (closed to .90) and RMSEA (closed to 0.60) are within an acceptable range. The combined GOF indices qualified the model for an appropriate fit of the data (Marsh et al., 2004). This finding is consistent with most reported literature on the validation of R-SPQ-2F (e.g., Socha & Sigler, 2014). The negative correlation (r = -.52, $p < .05$) found between deep and surface subscales is an indication of discriminant

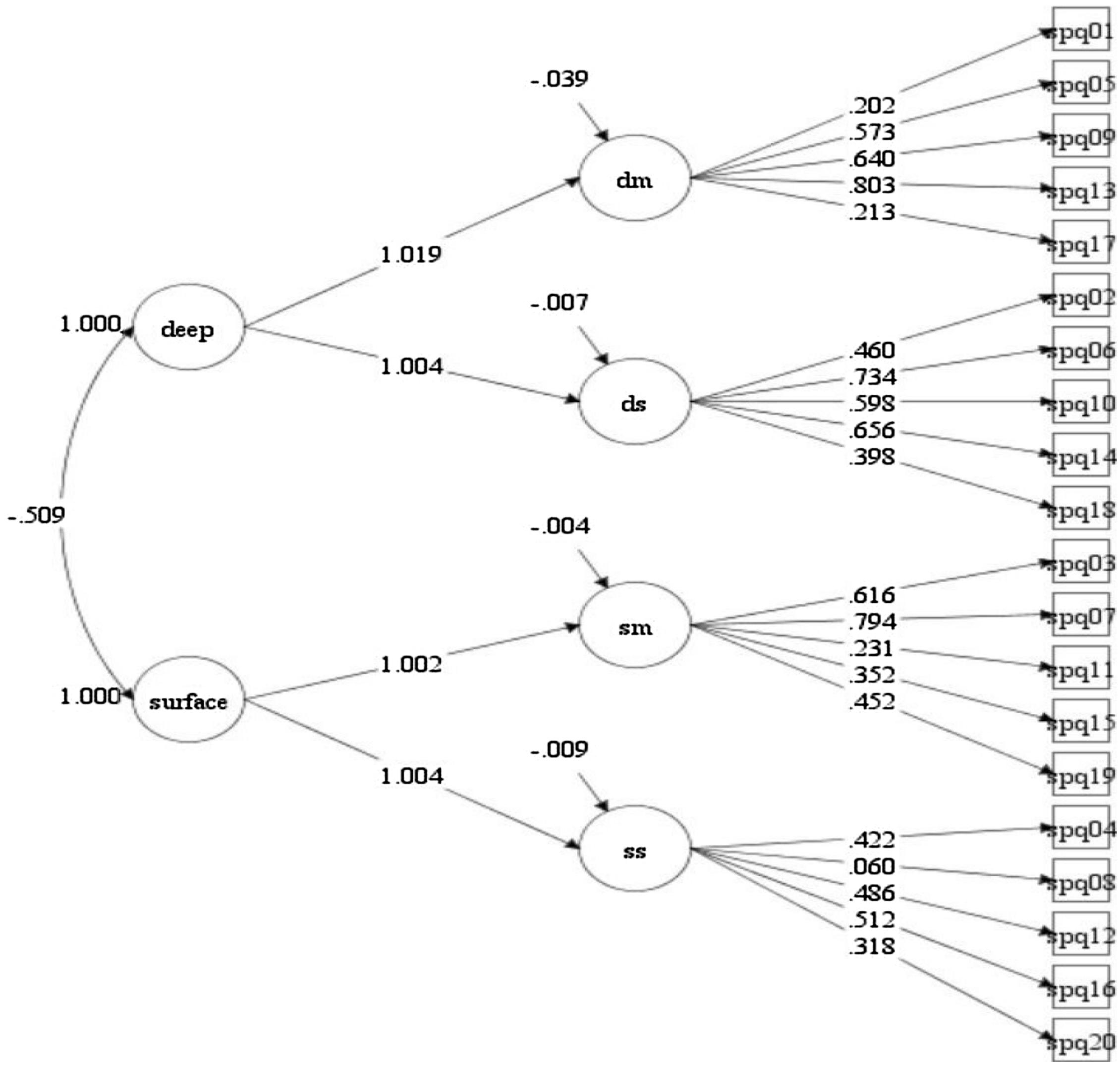


Fig. 3. Model 3 diagram with standard estimated parameters.

Table 5
Mplus output of model 4: Selected GOF statistics.

Tests of model fits		
Chi-Square Test of Model Fit	with item 8	without item 8
Value	522.179	377.676
Degrees of Freedom	169	151
p-value	0.0000	0.000
CFI/TLI		
CFI	0.769	0.844
TLI	0.740	0.824
Number of Free Parameters	101	96
RMSEA (Root Mean Square Error of Approximation)		
Estimate	0.091	0.077
90 Percent C.I.	0.082 0.100	0.067 0.087
Probability RMSEA < = .05	0.000	0.000
SRMR (Standardized Root Mean Square Residual)		
Value	0.081	0.072

validity between the subscales. This is expected and it is consistent with the literature (e.g., López-Aguado & Gutiérrez-Provecho, 2018).

4.5. Analysis of models 5 and 6

The one-factor models 5 and 6 containing 10 items on deep approach subscale and 9 items on surface approach subscale of R-SPQ-2F were fitted and reported in Table 6. The results show good model fits which confirmed the unidimensionality of each subscale. These results partly agreed with Biggs et al. (2001) and some other literature that have reported unidimensionality of items on each of deep and surface subscales (e.g., López-Aguado & Gutiérrez-Provecho, 2018).

4.6. Reliability of R-SPQ-2F

Reliability estimate of the whole R-SPQ-2F was checked as well as its subscales using latent variable approach suggested in (Raykov & Marcoulides, 2016). This approach has been proven to perform better than the conventional Cronbach's alpha estimates under violations of multiple assumptions like normality, skewness, etc. Simplified formulae adapted for the current research involving unidimensional and two-factor multidimensional scale are presented in Eqs. (1) and (2).

$$r_{RM} = \frac{(\sum_{i=1}^{n} L_i)^2}{(\sum_{i=1}^{n} L_i)^2 + \sum_{i=1}^{n} V_i} \quad (1)$$

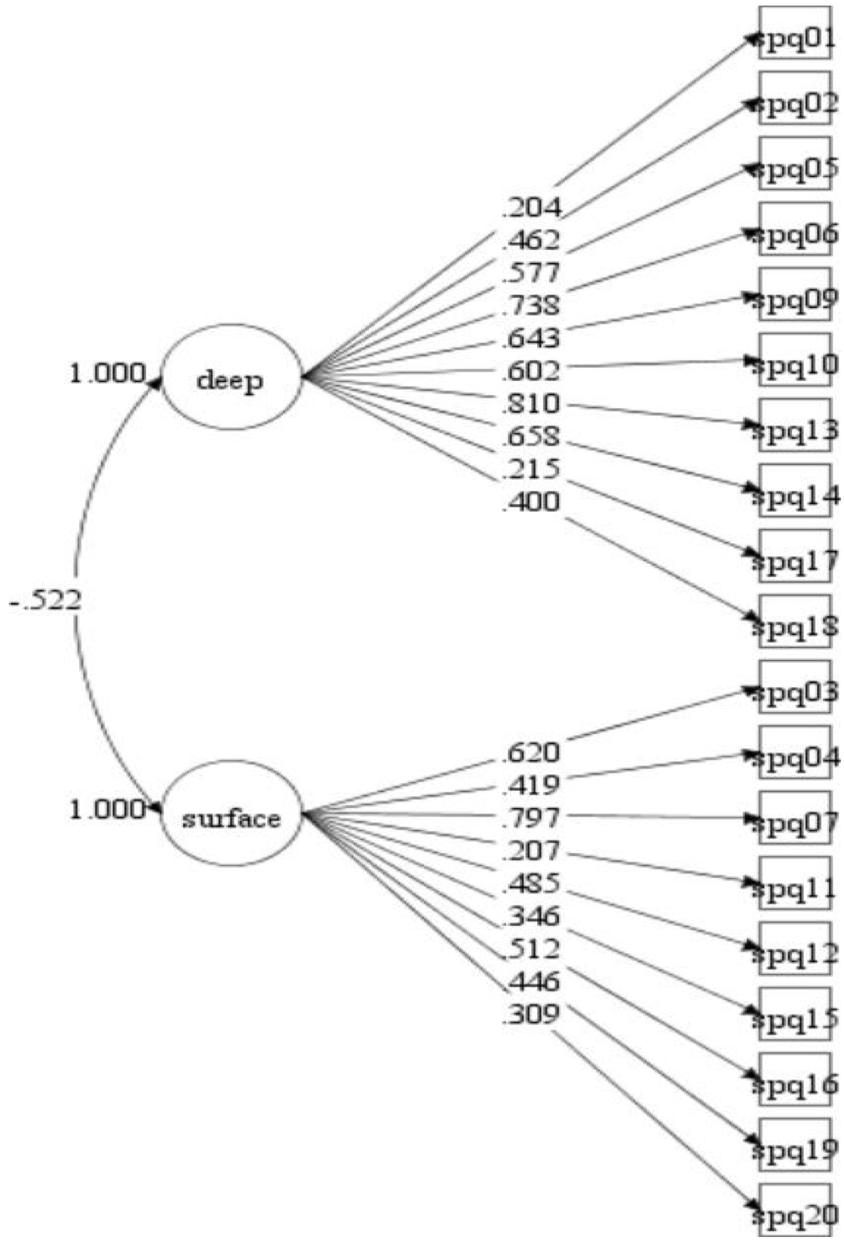


Fig. 4. Model 4 diagram with standard estimated parameters.

Table 6
Mplus output of models 5 and 6: Selected GOF statistics.

Tests of model fits		
Chi-Square Test of Model Fit	Model 5 (Deep)	Model 6 (Surface)
Value	92.884	65.58
Degrees of Freedom	35	46
p-value	0.0000	0.000
CFI/TLI		
CFI	0.943	0.919
TLI	0.926	0.887
Number of Free Parameters	50	96
RMSEA (Root Mean Square Error of Approximation)		
Estimate	0.081	0.078
90 Percent C.I.	0.061 0.101	0.054 0.101
Probability RMSEA < = .05	0.006	0.027
SRMR (Standardized Root Mean Square Residual)		
Value	0.047	0.050

$$r_{RM} = \frac{(\sum_{i=1}^{n} L_i)^2 + (\sum_{j=1}^{m} L_j)^2 + 2 * F_{12} *(\sum_{i=1}^{n} L_i)*(\sum_{j=1}^{m} L_j)}{(\sum_{i=1}^{n} L_i)^2 + (\sum_{j=1}^{m} L_j)^2 + 2 * F_{12} *(\sum_{i=1}^{n} L_i)*(\sum_{j=1}^{m} L_j) + \sum_{i=1}^{n} V_i + \sum_{j=1}^{m} V_j} \quad (2)$$

In Eqs. (1) and (2), r_{RM} is the Raykov and Marcoulides' correlation coefficient with a value ranges from 0 to 1 and interpreted like the Cronbach alfa coefficients with 0 to 1 indicative of item internal consistency from weakest to strongest (Raykov & Marcoulides, 2016). L_i' s and L_j' s are standardized factor loadings of subscale indicators, V_i' s and V_j' s are standardized unique variance computed by subtracting respective squared factor loading from 1 of each subscale indicator and F_{12} is the standardized covariance between factors 1 and 2. Using Eqs. (1) and (2) it was found that deep and surface subscales as well as the whole instrument have reliability coefficients of .81, .72 and .63 respectively. These are indicative of high reliability of the instrument. They are higher than the ones reported in (Biggs et al., 2001; López-

Aguado & Gutiérrez-Provecho, 2018) and within the ranges reported in other literature (e.g., Justicia et al., 2008; Socha & Sigler, 2014).

5. Conclusion

The cultural sensitivity of R-SPQ-2F has been attracting attention of educationists over a decade ago. Perhaps, this increased attention was prompted by the global quest for breeding university students towards deep approach to learning. This study was aimed at addressing issues related to the construct validation of this instrument as applied to the Norwegian context. In order to explain the factor structure of this instrument a series of confirmatory factor analyses were conducted, and several hypothesized models were evaluated. The best explanation found was a two-factor structure of the instrument measuring deep approach with 10 items (as theorized in the English version) and surface approach with 9 items (excluding item 8). The two-factor solution of the R-SPQ-2F found in the current study is in-line with a handful of adaptations of the instrument to Turkish (Önder & Besoluk, 2010), Spanish (Merino & Kumar, 2013), Chinese (Xie, 2014), etc.

There are several reasons that justify the removal of item 8 ("I learn some things by rote, going over and over them until I know them by heart even if I do not understand them") from the instrument. First, apart from its nonsignificant factor loading as revealed by CFA, a close look at the item itself raises some concerns. It includes some terms like "rote", "going over and over them" and "learning by heart" which seems confounding and could pose some levels of confusion to students (López-Aguado & Gutiérrez-Provecho, 2018). More so, a local misfit of this item as well as its nonsignificant factor loading have been reported in literature and its removal from the scale was recommended to obtain a valid measure (e.g., Immekus & Imbrie, 2010; Socha & Sigler, 2014).

The results, though partly contrary to the hypothesized models of Biggs et al. (2001) were similar to the ones in related studies (e.g., López-Aguado & Gutiérrez-Provecho, 2018; Socha & Sigler, 2014). The findings of the current study being the first of its kinds in Norway to the best of our knowledge have provided insights into the cultural sensitivity of the R-SPQ-2F. We acknowledge the study of Diseth (2001) on approaches to learning in the Norwegian context and the contributions made in relating approaches to learning with other constructs e.g. performance (Diseth, Pallesen, Brunborg, & Larsen, 2010). Some of our findings such as classification of learning approaches into deep and surface partly overlapped. However, their studies have relied on an old instrument, ASSIST, which was considered too lengthy and almost outdated as a revised version had been invoked. This current study made a significant shift form this old trend by considering a concise and easily interpretable measure of approaches to learning in the Norwegian context. Further, the studies of Diseth and colleagues (e.g., Diseth et al., 2010) have concentrated on Psychology students in contrary to engineering students which were the focus of this current study.

The approach adopted in this study has relied on recent development in structural equation modeling for psychometric studies and very selective in the choice of statistical tools. However, the results presented here are representative of the data which might not be generalizable to other cultural backgrounds. It is therefore recommended to make further explorations of this instrument before adapting it to another cultural context. A limitation acknowledged in this current study is the inability to investigate the measurement invariance of the proposed model across different groups. It is hoped that more validation studies on the hypothesized model in an independent sample and comparison of it with other models will be conducted. The instrument as attached in Appendix A will be indispensable to university teachers within Norway and the Mplus syntax codes provided in Appendix B could be modified for further related studies in this area. It is recommended that scoring should be done as proposed by Biggs et al. (2001) and scaled after summing by dividing scores on deep approach (1 + 2 + 5 + 6 + 8 + 9 + 12 + 13 + 16 + 17) by 10 and scores on surface approach (3 + 4 + 7 + 10 + 11 + 14 + 15 + 18 + 19) by 9. This is conjectured to enhance the interpretation of the scores.

Acknowledgements

We acknowledge the support of faculty of engineering and science, University of Agder, Kristiansand, Norway as well as MatRIC, Centre for Research, Innovation and Coordination of Mathematics Teaching for supporting this research. The anonymous reviewers are also acknowledged for their comments and recommendations towards the improvement of the article.

Appendix A. Revised Study Process Questionnaire (R-SPQ-2F)

This questionnaire has a number of questions about your attitudes towards mathematics and your usual way of studying it.

There is no *right* way of studying. It depends on what suits your own style and the course you are studying. It is accordingly important that you answer each question as honestly as you can.

Place the mark (✓) at the appropriate option to each statement. The letters alongside each number stand for the following response.

A—this item is *never* or *only rarely* true of me

B—this item is *sometimes* true of me

C—this item is true of me about *half the time*

D—this item is *frequently* true of me

E—this item is *always* or *almost always* true of me

	Statement on approaches to learning mathematics	A	B	C	D	E
1	I find that at times studying gives me a feeling of deep personal satisfaction.					
2	I find that I have to do enough work on a topic so that I can form my own conclusions before I am satisfied.					
3	My aim is to pass the course while doing as little work as possible.					
4	I only study seriously what's given out in class or in the course outlines.					
5	I feel that virtually any topic can be highly interesting once I get into it.					
6	I find most new topics interesting and often spend extra time trying to obtain more information about them					
7	I do not find my course very interesting so I keep my work to the minimum.					
8	I find that studying academic topics can at times be as exciting as a good novel or movie.					
9	I test myself on important topics until I understand them completely.					
10	I find I can get by in most assessments by memorising key sections rather than trying to understand them.					
11	I generally restrict my study to what is specifically set as I think it is unnecessary to do anything extra.					
12	I work hard at my studies because I find the material interesting.					
13	I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes.					
14	I find it is not helpful to study topics in depth. It confuses and wastes time, when all you need is a passing acquaintance with topics.					
15	I believe that lecturers shouldn't expect students to spend significant amounts of time studying material everyone knows won't be examined.					

16	I come to most classes with questions in mind that I want answering
17	I make a point of looking at most of the suggested readings that go with the lectures.
18	I see no point in learning material which is not likely to be in the examination.
19	I find the best way to pass examinations is to try to remember answers to likely questions.

Appendix B. Mplus syntax codes used for the analysis

```
TITLE: CFA MODEL 1 BIGGS et al., 2001.
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   CATEGORICAL ARE SPQ01-SPQ20;
   USEVARIABLES ARE SPQ01-SPQ20;
   MISSING ARE ALL (-1);
ANALYSIS:
ESTIMATOR IS WLSMV;
MODEL:
DM by SPQ01 SPQ05 SPQ09 SPQ13 SPQ17;
DS by SPQ02 SPQ06 SPQ10 SPQ14 SPQ18;
SM by SPQ03 SPQ07 SPQ11 SPQ15 SPQ19;
SS by SPQ04 SPQ08 SPQ12 SPQ16 SPQ20;
DM WITH SS@0;
DS WITH SS@0;
DS WITH SM@0;

OUTPUT: MODINDICES STANDARDIZED TECH4;
```

```
TITLE: CFA MODEL 2 BIGGS et al. 2001
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE DM DS SM SS;
   MISSING ARE ALL (-1);
ANALYSIS:
ESTIMATOR IS ML;
MODEL:
Deep by DM DS;
Surface by SM SS;

OUTPUT: MODINDICES STANDARDIZED TECH4;
```

```
TITLE: CFA HIERARCHICAL MODEL 3
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   CATEGORICAL ARE SPQ01-SPQ20;
   USEVARIABLES ARE SPQ01-SPQ20;
   MISSING ARE ALL (-1);
ANALYSIS:
ESTIMATOR IS WLSMV;
MODEL:
DM by SPQ01 SPQ05 SPQ09 SPQ13 SPQ17;
DS by SPQ02 SPQ06 SPQ10 SPQ14 SPQ18;
SM by SPQ03 SPQ07 SPQ11 SPQ15 SPQ19;
SS by SPQ04 SPQ08 SPQ12 SPQ16 SPQ20;
Deep by DM* DS;
Surface by SM* SS;
DM DS(1);
SM SS(1);
Deep@1;
Surface@1;

OUTPUT: STANDARDIZED TECH4;
```

```
TITLE: CFA 2-factor model without item 8
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE SPQ01-SPQ06 SPQ07 SPQ09-SPQ20;
   CATEGORICAL ARE SPQ01-SPQ20;
   MISSING ARE ALL (-1);

ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION = THETA;
ITERATION = 1000;
MODEL:
Deep by SPQ01 SPQ02 SPQ05 SPQ06 SPQ09 SPQ10 SPQ13
SPQ14 SPQ17 SPQ18;
Surface by SPQ03 SPQ04 SPQ07 SPQ11 SPQ12 SPQ15 SPQ16
SPQ19 SPQ20;

OUTPUT: MODINDICES (ALL) STDYX TECH4;
```

```
TITLE: CFA unidimentionality deep approach
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE SPQ01 SPQ02 SPQ05 SPQ06 SPQ09
SPQ10 SPQ13 SPQ14 SPQ17 SPQ18;
   CATEGORICAL ARE SPQ01 SPQ02 SPQ05 SPQ06 SPQ09
SPQ10 SPQ13 SPQ14 SPQ17 SPQ18;
   MISSING ARE ALL (-1);

ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION = THETA;
ITERATION = 10000;
MODEL:
Deep by SPQ01 SPQ02 SPQ05 SPQ06 SPQ09 SPQ10 SPQ13
SPQ14 SPQ17 SPQ18;

OUTPUT: STANDARDIZED STDY TECH4;
```

```
TITLE: CFA unidimentionality surface
DATA:
FILE IS "C:\Users\SPQ.dat";
VARIABLE:
NAMES ARE SPQ01-SPQ20 DA SA DM DS SM SS;
   USEVARIABLES ARE SPQ03 SPQ04 SPQ07 SPQ11 SPQ12
SPQ15 SPQ16 SPQ19 SPQ20;
   CATEGORICAL ARE SPQ03 SPQ04 SPQ07 SPQ11 SPQ12
SPQ15 SPQ16 SPQ19 SPQ20;
   MISSING ARE ALL (-1);
 ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION = THETA;
ITERATION = 10000;
MODEL:
Surface by SPQ03 SPQ04 SPQ07 SPQ11 SPQ12 SPQ15 SPQ16
SPQ19 SPQ20;
SPQ03 WITH SPQ07;

OUTPUT: STANDARDIZED STDY TECH4;
```

References

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243–260. https://doi.org/10.1016/j.edurev.2010.06.001.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.

Biggs, J. B. (2012). What the student does: Teaching for enhanced learning. *Higher Education Research & Development, 31*(1), 39–55. https://doi.org/10.1080/07294360.2012.642839.

Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two factor study processquestionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*, 133–149.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research.* New York, London: The Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258. https://doi.org/10.1177/0049124192021002005.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* New York: Routledge, Taylor & Francis Group.

Chan, Y. M., & Sheung Chan, C. M. (2010). Approach to learning of sub-degree students in Hong Kong. *Research in Education, 84*(1), 65–78.

Diseth, Å. (2001). Validation of a Norwegian version of the approaches and study skills inventory for students (ASSIST): Application of structural equation modelling. *Scandinavian Journal of Educational Research, 45*(4), 381–394.

Diseth, Å., Pallesen, S., Brunborg, G. S., & Larsen, S. (2010). Academic achievement among first semester undergraduate psychology students: The role of course experience, effort, motives and learning strategies. *Higher Education, 59*, 335–352.

Fryer, L. K., Ginns, P., Walker, R. A., & Nakao, K. (2012). The adaptation and validation of the CEQ and the R-SPQ-2F to the Japanese tertiary environment. *The British Journal of Educational Psychology, 82*(Pt 4), 549–563. https://doi.org/10.1111/j.2044-8279.2011.02045.x.

Fryer, L. K., & Vermunt, J. D. (2018). Regulating approaches to learning: Testing learning strategy convergences across a year at university. *The British Journal of Educational Psychology, 88*(1), 21–41. https://doi.org/10.1111/bjep.12169.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research, 41*(1), 65–83. https://doi.org/10.1207/s15327906mbr4101_5.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Immekus, J. C., & Imbrie, P. K. (2010). A test and cross-validation of the revised two-factor study process questionnaire factor structure among western university students. *Educational and Psychological Measurement, 70*(3), 495–510. https://doi.org/10.1177/0013164409355685.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: Guide to the program and applications* (2nd ed.). Chicago, IL: SPSS.

Justicia, F., Pichardo, M. C., Cano, F., Berbén, A. B. G., & De la Fuente, J. (2008). The revised two-factor study process questionnaire (R-SPQ-2F): Exploratory and confirmatory factor analyses at item level. *European Journal of Psychology of Education, 23*(3), 355–372.

Kember, D. (1990). Cultural specificity of approaches to study. *The British Journal of Educational Psychology, 60*, 356–363.

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances. *Sociological Methods & Research, 41*(1), 124–167. https://doi.org/10.1177/0049124112442138.

Lake, W. W., Boyd, W. E., & Boyd, W. (2017). Understanding how students study: The genealogy and conceptual basis of a widely used pedagogical research tool, Biggs' study process questionnaire. *International Education Studies, 10*(5), 100–108. https://doi.org/10.5539/ies.v10n5p100.

Leung, D. Y. P., Ginns, P., & Kember, D. (2008). Examining the cultural specificity of approaches to learning in universities in Hong Kong and Sydney. *Journal of Cross-cultural Psychology, 39*(3), 251–266. https://doi.org/10.1177/0022022107313905.

López-Aguado, M., & Gutiérrez-Provecho, L. (2018). Checking the underlying structure of R-SPQ-2F using covariance structure analysis / Comprobación de la estructura subyacente del R-SPQ-2F mediante análisis de estructura de covarianza. *Cultura y Educación/ Culture and Education, 30*(1), 105–141. https://doi.org/10.1080/11356405.2017.1416787.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Maciejewski, W., & Merchant, S. (2016). Mathematical tasks, study approaches, and course grades in undergraduate mathematics: A year-by-year analysis. *International Journal of Mathematical Education in Science and Technology, 47*(3), 373–387. https://doi.org/10.1080/0020739x.2015.1072881.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling A Multidisciplinary Journal, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2.

Marton, F., & Säljö, R. (1976a). On qualitative differences in learning - II: Outcome as a function of the learner's conception of the task. *The British Journal of Educational Psychology, 46*, 115–127.

Marton, F., & Säljö, R. (1976b). On qualitative differences in learning I: Outcome and process. *The British Journal of Educational Psychology, 46*, 4–11.

Merino, S. C., & Kumar, P. R. (2013). Validación estructural del R-SPQ-2F: Un análisis factorial confirmatorio. *Revista Digital de Investigación en Docencia Universitaria, 7*, 113–127.

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus user's guide. (Eighth ed.). Los Angeles, CA: Muthén & Muthén.

Önder, I., & Besoluk, S. (2010). Adaptation of revised two factor study process questionnaire (R-SPQ-2F) to Turkish. *Education and Science, 35*(157), 55–67.

Raykov, T., & Marcoulides, G. A. (2016). Scale reliability evaluation under multiple assumption violations. *Structural Equation Modeling A Multidisciplinary Journal, 23*(2), 302–313. https://doi.org/10.1080/10705511.2014.938597.

Shaik, S. A., Almarzuqi, A., Almogheer, R., Alharbi, O., Jalal, A., & Alorainy, M. (2017). Assessing Saudi medical students learning approach using the revised two-factor study process questionnaire. *International Journal of Medical Education, 8*, 292–296. https://doi.org/10.5116/ijme.5974.7a06.

Socha, A., & Sigler, E. A. (2014). Exploring and "reconciling" the factor structure for the revised two-factor study process questionnaire. *Learning and Individual Differences, 31*, 43–50. https://doi.org/10.1016/j.lindif.2013.12.010.

Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) handout. *Structural Equation Modeling A Multidisciplinary Journal, 23*(6), 777–781. https://doi.org/10.1080/10705511.2016.1217487.

Stes, A., De Maeyer, S., & Van Petegem, P. (2013). Examining the cross-cultural sensitivity of the revised two-factor study process questionnaire (R-SPQ-2F) and validation of a Dutch version. *PloS One, 8*(1), e54099. https://doi.org/10.1371/journal.pone.0054099.

Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling A Multidisciplinary Journal, 22*(4), 568–580. https://doi.org/10.1080/10705511.2014.937669.

Tait, H., Entwistle, N. J., & McCune, V. (1998). ASSIST: A re-conceptualisation of the approaches to studying inventory. In C. Rust (Ed.). *Improving students as learners* (pp. 262–271). Oxford: Oxford Brookes University.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

Valadas, S., Gonçalves, F., & Faísca, L. (2010). Approaches to studying in higher education Portuguese students: A Portuguese version of the approaches and study skills inventory for students. *Higher Education, 59*, 259–275.

Weller, J. M., Henning, M., Civil, N., Lavery, L., Boyd, M. J., & Jolly, B. (2013). Approaches to learning for the ANZCA final examination and validation of the revised study process questionnaire in specialist medical training. *Anaesthesia and Intensive Care, 41*, 631–640.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913–934. https://doi.org/10.1177/0013164413495237.

Xie, Q. (2014). Validating the revised two-factor study process questionnaire among Chinese university students. *The International Journal of Educational and Psychological Assessment, 16*(1).

Paper II

Zakariya, Y. F. (2019). Study approaches in higher education mathematics: Investigating the statistical behaviour of an instrument translated into Norwegian. *Education Sciences, 9*(3), 191. doi:10.3390/educsci9030191 *education sciences*

MDPI

Article

Study Approaches in Higher Education Mathematics: Investigating the Statistical Behaviour of an Instrument Translated into Norwegian

Yusuf F. Zakariya

Department of mathematical sciences, University of Agder, 4630 Kristiansand, Norway; yusuf.zakariya@uia.no

Received: 31 May 2019; Accepted: 19 July 2019; Published: 20 July 2019

check for updates

Abstract: The revised two-factor study process questionnaire (R-SPQ-2F) has widely been considered valid and reliable in many contexts for measuring students' learning approaches. However, its cultural specificity has generated considerable discussion, with inconclusive results when translated to different languages. This paper provides more insights into the construct validity of a Norwegian version of this instrument. The R-SPQ-2F is composed of ten items designed to expose deep learning approaches and 10 items designed to expose surface learning approaches. A survey research design involving a sample of 253 first year university students in a mathematics course was adopted. Ten hypothesized models were compared using a series of confirmatory factor analyses following the model proposals reported in the literature. A weighted least square mean and variance adjusted (WLSMV) estimator was used to enhance model parameter estimations under multiple violations of assumptions inherent in ordinal data. The results favored a two first-order factor model with ten items measuring the deep approach and nine items measuring the surface approach including a deletion of one item from this instrument. The findings of this study provide empirical evidence for the cultural specificity of the instrument that is consistent with the literature. The R-SPQ-2F is therefore recommended to assess students' approaches to learning, and further studies into its cultural specificity are recommended.

Keywords: university mathematics; deep learning; surface learning; multivariate statistics; confirmatory factor analysis

1. Introduction

Empirical evidence has shown that students' learning approaches contribute significantly to their academic success in higher education (e.g., [1,2]). Learning approaches could be conceived as an individual's adopted predispositions when dealing with tasks and strategies used to process learning materials which can be deep or surface in nature [3,4]. A deep approach to learning involves concentration on latent meanings of the material to be learned, while a surface approach entails memorization and less priority on the conveyed messages in the presented tasks. Deep learning has been an emerging focus of higher education studies in preparing future leaders for our ever-increasingly diverse society [2]. For many decades, educators have been challenged by the proper conceptualization and operationalization of students' approaches to learning (SAL).

The SAL theory of Marton and Säljö [5,6] uses phenomenography coupled with some constructivist perspectives of Biggs [7,8] and has provided theoretical frameworks for conceptualizing students' approaches to learning. This is evident in the way approaches to learning have been defined to include motives, predispositions, styles, strategies used in adopting a process of learning tasks. Moreover, the classification of approaches students adopted when learning into 'surface' and 'deep' has greatly influenced SAL measuring instruments. A widely studied instrument for measuring

SAL is the Biggs' [9] study process questionnaires [10]. This instrument has undergone several revisions and validations from its initial 72-item to the present 20-item two-factor revised study process questionnaire [9]. It has gained equally wide acceptance among educators, with many studies on its psychometric properties, and Cronbach's alpha ranges 0.57–0.85 have been reported as evidence of the item's internal consistency [9,11].

However, the cultural specificity of the two-factor revised study process questionnaires (R-SPQ-2F) has generated considerable discussion with inconclusive results when the instrument has been translated to different languages (e.g., [12,13]). Apart from the two models hypothesized by Biggs, Kember and Leung [9], several alternative models have been proposed, and some items were deleted to achieve modest fits in explaining the underlying factor structures of the instrument. This current study was framed with the sole aim of exploring and comparing alternative models that best explain the construct validity of the R-SPQ-2F when translated to the Norwegian context. This article is a continuation of a work reported in [14], where the Biggs' et al. [9] hypothesized models were investigated and found to poorly represent Norwegian data with non-admissible solutions. In the earlier work, a new model for the R-SPQ-2F was proposed and confirmed, and the model fits and scale reliability were investigated and reported. The purpose of this paper is to contribute to this body of research and expose some observable methodological weaknesses inherent in some reported hypothesized models in literature.

2. Literature Review

Studies on the cultural specificity of the R-SPQ-2F can broadly be classified into two major categories. The first category represents those that report first-order two-factor structures—the deep approach (DA) and the surface approach (SA)—as the best explanatory models for the instrument with ten items on each subscale [15–18]. This first category can further be divided into those that include error covariance—the presence of a systematic commonly shared variance—between indicators (e.g., [17]) and those that did not include the covariance (e.g., [16]). However, Biggs et al. [9] were the first to start a discussion on the factor structure of their then newly developed instrument, the R-SPQ-2F, by hypothesizing and testing two models. The first model was a first-order four-factor model containing 'deep motive,' 'deep strategy,' 'surface motive,' and 'surface strategy' measured by five items each. The first model was tested and found to fit their 495 data with a comparative fit index (CFI) of 0.904 and a standardized root mean squared residual (SRMR) of 0.058 [9]. Further, CFIs of 0.997, 0.998, 0.988, and 0.998 and SRMRs of 0.01, 0.02, 0.02 and 0.02 were also reported on 'deep motive,' 'deep strategy,' 'surface motive,' and 'surface strategy' subscales, respectively. The second model was a first-order two-factor model containing deep and surface approaches with two indicators each—motive and strategy—gotten by corresponding item parceling (adding scores on five items) in the first model. The results of the second model also suggest a good model fit with a CFI of 0.992 and an SRMR of 0.015, both of which are within the proposed cutoffs by Hu and Bentler [19].

These two hypothesized models of Biggs et al. [9] have steered heated debates among educators and methodologists when subjected to confirmatory analysis in an independent cultural context. For example, the two models were tested and found to fairly explain the factor structure of the R-SPQ-2F when translated to Spanish in a study involving 836 undergraduate students, out of which 314 were used for exploratory factor analysis and the remaining 522 were used for confirmatory factor analysis [15]. An alternative model of a first-order two-factor model was proposed and tested containing the deep and surface approaches measured by their corresponding ten items each as theorized in [9]. The results suggest a modest fit with a significant χ^2-value (169) = 645.77, $p < 0.05$, goodness of fit index (GFI) = 0.95, SRMR = 0.09, root mean square error of approximation (RMSEA) = 0.07, non-normed fit index (NNFI) = 0.91, CFI = 0.92, parsimony normed fit index (PNFI) = 0.80, and parsimony goodness of fit index (PGFI) = 0.76. In another study, Önder and Besoluk [18] reported a Turkish validation of the instrument when administered to 528 undergraduate students. Their findings also identified a first-order two-factor model as the best explanation for the construct validity of the R-SPQ-2F. Their results involved a significant χ^2-value (166) = 487.95, $p < 0.05$, GFI = 0.89, SRMR = 0.07, RMSEA = 0.06, NNFI = 0.90, CFI = 0.93, PGFI = 0.92, incremental fit index (IFI) = 0.93, and relative fit index (RFI) = 0.88. A major difference between these results and that of Justicia, Pichardo, Cano, Berbén and De la Fuente [15] was the inclusion of error covariance between items 8 and 10 as well as between items 11 and 20.

Non-admissible solutions and poor fits for the hypothesized models in [9] were also reported in a study involving 269 university and non-university students [17]. Following confirmatory factor analysis results, a first-order two-factor model was identified as the best explanation for the construct validity of the R-SPQ-2F. A significant χ^2-value (168) = 259.32, $p < 0.05$, was also reported and coupled with SRMR = 0.07, RMSEA = 0.05, Tucker-Lewis index (TLI) = 0.95, and CFI = 0.96. Similar to the findings of Önder and Besoluk [18] an error covariance was also defined between items 4 and 14 in order to achieve a good model fit. Corroborative results can also be found in a Chinese validation of the R-SPQ-2F involving 439 university students, in which a first-order two-factor model was also reported [16]. Table 1 presents a juxtaposition of the findings of these studies for easy comparison.

Table 1. Summary of findings on the first-order two-factor model of the revised two-factor study process questionnaire (R-SPQ-2F).

	[M1]-Justicia et al. [15]	**[M2]-Önder and Besoluk [18]**	**[M3]-Merino and Kumar [17]**	**[M4]-Xie [16]**
Error Covar.		8 and 10, 11 and 20	4 and 14	
Cor. DA/SA	−0.39	−0.51	−0.33	−0.35
df	169	166	168	169
χ^2-value	645.77	487.95	259.32	489.40
GFI	0.95	0.89		
SRMR	0.09	0.07	0.07	
RMSEA	0.07	0.06	0.05	0.07
NNFI/TLI	0.91	0.90	0.95	0.91
CFI	0.92	0.93	0.96	0.92
PNFI	0.80			0.79
PGFI	0.76	0.92		0.72
IFI		0.93		
RFI		0.88		

There seems to be a consistency in the results of previous studies presented in Table 1. They corroborate the theoretical explanation of indicators measuring the DA and the SA as proposed in [9] with an exclusion of additional subdivisions of each factor into motive and strategy. The negative standard correlation coefficients found in all the studies between the deep approach and the surface approach subscales are indicative of discriminant validity. A close look at the results of Merino and Kumar [17] as well as Önder and Besoluk [18] suggests a better fit of their models as compared to others. This can be deduced from their reduced χ^2-values and RMSEA within the range suggested in [19,20]. However, the inclusion of error covariance between some indicators in their models could pose some complications in the application and interpretation of the scale item scores by classroom teachers.

The second broad category of studies on the cultural specificity of the R-SPQ-2F are the reports of two first-order and four first-order factor structures with some items deleted to achieve good fits (e.g., [10,21]). The number of items deleted ranged from 2–5. Immekus and Imbrie [22] after establishing non-admissible solutions of the hypothesized Biggs' et al. [9] model, subjected the data from their first cohort of 1490 university students to an exploratory factor analysis (EFA). The results of their EFA gave four extracted latent factors which they identified as 'deep motive,' 'deep strategy,' 'surface motive,' and 'surface strategy' after rotating using Promax. Five items that exhibit substantial cross-loading were removed from the model. The first-order four-factor model was then subjected to a confirmatory factor analysis in an independent cohort of 1533 university students' sample. The results of a confirmatory analysis suggest a modest fit with a significant χ^2-value (114) = 568.54, $p < 0.05$, RMSEA = 0.05, and CFI = 0.96. Surprisingly, relatively high positive correlations of 0.76 and 0.59 were found between 'deep motive' and 'deep strategy' as well as 'surface motive' and 'surface strategy,' respectively.

No empirical evidence was found to support the first Biggs' et al. model in the Japanese validation of the R-SPQ-2F reported in [21]. However, a modest fit for the second Biggs' et al. model involving a first-order two-factor model with item parceling on each deep and surface approach latent factors was confirmed. The study involved 269 university students distributed across different programs in a Japanese tertiary institution. The results of their confirmatory analysis did not include the χ^2-value, instead an RMSEA = 0, CFI = 1, and TLI = 1 coupled with a positive correlation cooeficient of 0.30 between the deep and surface approach latent factors were reported [21]. There are some reservations with respect to these results. First, the goodness of fits (GOF) indices indicate a perfect fit of the model, which appears to be unrealistic. However, an observed methodological issue could stem from the degress of freedom (though not reported), which is 1. This could make it difficult for the variance/covariance matrix to be positively definite. Unfortunately, nothing was mentioned in the article with respect to this matrix. Another methodological difficulty that could even lead to the rejection of this model is the positive correlation of 0.30 reported between deep and surface approaches. This shows a non-discriminating capacity of this model between the deep and surface approaches which is contrary to both the theoretical and the conceptual interpretations of the instrument.

More so, Stes, De Maeyer and Van Petegem [12] could not also find any supportive empirical evidence for both models hypothesized by Biggs et al. [9] in the validation of their Dutch version of the instrument involving 1974 effetive sample of students distributed across diverse university programs. For this reason, an exploratory factor analysis was conducted on a randomly selected 963 cases from the total sample, using maximum likelihood for factor extraction and an oblique rotation. Five factors were initially identified, and these were later collapsed into four factors—study is interesting (SI), spending extra time (ST), minimal effort (ME), and learning by heart (LH)—after a series of confirmatory factor analyses and item deletions. The final fitted solution was a first-order four-factor model with three items measuring SI, four items measuring ST, five items measuring ME and three items measuring LH. The final chi-squared statitistic as well as the degree of freedom were not reported. However, some GOF indices such as GFI = 0.95, absolute goodness of fit index (AGFI) = 0.93, RMSEA = 0.06, CFI = 0.94, and PGFI = 0.66 were reported as evidence of a good fit for their model. Further, relatively high correlation coefficinets of 0.76 and 0.62 were found between SI and ST as well as between ME and LH, respectively.

In an attempt to reconcile between variant inconclusive models results on the R-SPS-2F, Socha and Sigler [13] conducted a validation study on the instrument invoving 868 university students. In their study, eight models were compared using a confirmatory factor analysis, and a first-order two-factor solution was found as the best explanation for the construct validity of the instrument involving the deletion of two items from the original version. Their final results included a significant χ^2-value (134) = 504.83, $p < 0.05$, SRMR = 0.05, RMSEA = 0.06, and CFI = 0.95 and a negative correlation of −0.38 between the deep and surface approach latent factors. Similar results can also be found in another Spanish validation of the R-SPQ-2F involving 279 university students, in which a first-order two-factor model coupled with two item deletion was also reported [10]. Table 2 presents a juxtaposition of the findings of these studies for easy comparison.

Table 2. Summary of findings on the two-factor and four-factor models of the R-SPQ-2F.

	[M5]-Immekus and Imbrie [22]	[M6]-Fryer et al. [21]	[M7]-Stes et al. [12]	[M8]-Socha and Sigler [13]	[M9]-López-Aguado and Gutiérrez-Provecho [10]
Item deleted	1, 3, 7, 13, 15		2, 3, 7, 10, 17	7, 8	7, 8
Cor. DA/SA	DM/DS = 0.76 SS/SM = 0.59	0.30	SI/ST = 0.76 ME/LH = 0.62	−0.38	−0.41
df	114			134	134
χ^2-value	568.54			504.83	226.53
GFI			0.95		0.91
AGFI			0.93		
SRMR				00.05	0.05
RMSEA	0.05	0	0.06	0.06	0.05
CFI	0.96	1	0.94	0.95	0.91
TLI		1			0.90
PGFI			0.66		0.72

Note: DM = Deep motive; DS = Deep strategy; SM = Surface motive; SS = Surface strategy; SI = Study is interesting; ST = Spending extra time; ME = Minimal effort; and LH = Learning by heart.

The results presented in Table 2 reveal variant and inconclusive solutions of the models. These can be ascribed to some methodological issues inherent in the factor analysis procedure as well as the cultural sensitivity of the instrument. For example, Immekus and Imbrie [22], after establishing non-admissible solutions of the hypothesized models in [9], subjected their data to an exploratory factor analysis (FA). Difficulties arose when some indicators loaded (loadings greater than |0.3|) on more than one extracted factor. Rather than seeking theoretical explanations for this observation, they opted to delete these indicators from the scales. For instance, item 1 loaded on deep motive (DM) and deep strategy (DS) with 0.31 and 0.42 oblique rotated loadings, respectively. This could be suggestive of over-factoring in the extraction, especially when this item has been theorized to measure both DM and DS. To support this claim, the high positive correlations of 0.76 and 0.59 reported between DM and DS as well as surface strategy (SS) and surface motives (SM), respectively, are indications of multicollinearity, which could be addressed by collapsing the subcategories.

A similar methodological issue is also perceived in the analysis of Stes et al. [12] with high positive correlations of 0.76 and 0.62 between SI and ST as well as ME and LH, respectively. Another methodological issue involved in the analysis of Stes et al. [12] and Fryer et al. [21] is the use of maximum likelihood estimator, which has been found to perform poorly in the analysis of ordinal data (e.g., [23,24]). It is also important to remark that SI combined with ST and LH combined with ME are other ways to refer to the DA and the SA, respectively. Later studies (e.g., [10]) seem to address some of these methodological issues, yet the cultural specificity of the R-SPQ-2F still remains an important consideration when adapted to a different language from English. Therefore, the current study sought to build on this literature in searching and evaluating hypothesized models to explain the construct validity of the R-SQP-2F in the Norwegian context.

3. Methods

3.1. Research Design and Materials

An appropriate design for this study is a survey type which can be used to justify the collection of data from a large number of students using questionnaires. A Norwegian version of the R-SPQ-2F was prepared using translation and back-translation approaches by two associate professors and a full professor of mathematics education in the research team. A comparison was made between translated versions, and some adjustments for language differences were done before the final version was prepared in an electronic form.

3.2. Sample and Data Collection Procedure

The sample comprised 253 undergraduate students on engineering programs in their first year across two Norwegian government universities. This was made of up of males (168) and females (72), including 13 students who did not respond to the gender item on the instrument, and their sample had a mean age of 19–22 years. The sample size was considered appropriate with a justification from reported computer simulated studies on the adequacy of a confirmatory factor analysis (CFA) sample size (e.g., [25]). The data were collected using an online questionnaire development kit (SurveyXact) and distributed to students (via university mails) who gave their consent to partake in the project. Class visits were organized to encourage students' participation, and some students completed a paper version of the questionnaires during the visits. The data were collected within 3 weeks, screened for outliers, and prepared for CFA. Very few data were missing at random and posed no concern for the analysis.

3.3. Data Analysis

CFA was used to test ten models with a weighted least square mean and variance adjusted (WLSMV) estimator using Mplus version 8.2 [26] for the analysis. The use of WLSMV was not accidental. It was because of its robust ability to perform better than maximum likelihood (ML), unweighted least squares (ULS), etc., in the analysis of ordinal data that violate multiple assumptions (see [23]). Model fits were assessed based on χ^2-values and combinations of the goodness of fit indices with some relaxations in the suggested cut-offs (CFI/TLI closed to or ≥ 0.9, SRMR ≤ 0.8, and RMSEA ≤ 0.60) proposed by Hu and Bentler [19]. This was necessary because of the type of data (ordinal) as well as a different estimator (WLSMV) as compared to the continuous data and ML estimator used in some simulation studies on cut-off criteria (e.g., [19,27]). The emerging results and discussion are presented in the next few paragraphs.

4. Results

The first set of results as presented in Table 3 represent the tested and hypothesized two first-order factor models of the R-SPQ-2F, as in the literature. Analyzed results from hypothesized model of Xie [16] are included in Table 3, and those of Justicia et al. [15] were excluded, because they both practically advocated the same model and the former is more recent. Notations and abbreviations used in Table 1 are repeated in Table 3, with M2 used for Önder and Besoluk [18], M3 used for Merino and Kumar [17], and M4 used for Xie [16].

Table 3. Selected CFA results of the two first-order factor hypothesized model of the R-SPQ-2F.

	M2	M3	M4
Error Covar.	8 and 10, 11 and 20	4 and 14	
Cor. DA/SA	−0.519	−0.507	−0.512
df	167	168	169
χ^2-value	495.212	517.980	522.179
p-value	0.000	0.000	0.000
CFI	0.785	0.771	0.769
TLI	0.756	0.741	0.740
SRMR	0.078	0.081	0.081
RMSEA	0.088	0.091	0.091
90% C.I.	0.079 0.097	0.082 0.100	0.082 0.100

The results presented in Table 3 show admissible solutions of the two first-order factor model of the R-SPQ-2F. Negative standard correlations found between deep and surface components are suggestive of a discriminant validity between these subscales. This could be interpreted to mean a student with a high score on deep approach items had a low score on surface approach items and vice versa. This makes sense and is conceptually sound. However, the high χ^2-values (496.21–522.18) coupled with out of range fit indices are indicative of the poor fit of these models. The model proposed by Önder and Besoluk [18] seems to perform better than others, with the lowest χ^2-value (167, $p < 0.05$) = 495.21 and RMSEA ≤ 0.08. Meanwhile, the two error covariances involved between item 8 and item 10, as well as item 11 and item 20, could pose some complications in classroom conceptual understanding and interpretation of scores from this instrument. Therefore, all these models are not statistically and conceptually fit to justify the construct validity of the R-SPQ-2F in the Norwegian context.

The second set of results concern the CFA of the two-factor and four-factor models hypothesized to explain construct validity of the R-SPQ-2F. The model result of Fryer et al. [21] was not included in Table 4 because it has been reported in [14]. The analyzed result showed a non-admissible solution of the model with a negative error variance on the SM indicator, a result that is suggestive of over-factoring in the model [14]. Further, analyzed results from the model hypothesized by López-Aguado and Gutiérrez-Provecho [10] were included in Table 4, but those of Socha and Sigler [13] were omitted because they are both practically advocating same model and the former is more recent. Notations and abbreviations used in Table 2 are repeated in Table 4 with M5 used for Immekus and Imbrie [22]. Mod. M5 was used for modified M5, M7 was used for Stes et al. [12], Mod. M7 was used for modified M7, and M9 was used for López-Aguado and Gutiérrez-Provecho [10].

Table 4. Selected CFA results of the two-factor and four-factor hypothesized models on the R-SPQ-2F.

	M5	Mod. M5	M7	Mod. M7	M9
Cor. DA/SA	DM/DS = 0.737 SS/SM = 0.366	−0.290	SI/ST = 1.020 ME/LH = 0.401	−0.284	−0.40
df	84	89	85	89	134
χ^2-value	152.278	289.254	138.318	257.024	301.440
p-value	0.000	0.000	0.0002	0.000	0.000
CFI	0.925	0.780	0.950	0.844	0.869
TLI	0.906	0.741	0.939	0.816	0.850
SRMR	0.054	0.077	0.052	0.072	0.068
RMSEA	0.057	0.094	0.050	0.086	0.070
90% C.I.	0.042 0.071	0.082 0.107	0.034 0.065	0.074 0.099	0.060 0.081

There seems to be indications of good fits in all the models analyzed and reported in Table 4. The reduced χ^2-values between 152.28 and 301.44 coupled with within suggested range indices may prompt one to conclude that M5 and M7 have been demonstrated as the best models. However, there was evidence of a gross misspecification and a high multicollinearity between DM and DS, which are suggestive of over-factoring in M5. This is evident with a high standardized correlation coefficient (r = 0.74, $p < 0.05$) between the DM and DS latent factors. This posed some methodological difficulties involved in trying to balance both the theoretical and conceptual understanding that could yield a substantive interpretation of scores from the instrument. Therefore, an attempt was made to revive this model as reported under the heading modified model 5 (Mod. M5). Here, items measuring DM and DS were merged to form a factor (DA), and those measuring SS and SM were merged to form another factor (SA). The resulting two-factor model was subjected to CFA, and selected GOF indices are presented in Table 4 with the heading Mod. M5. The χ^2-value (89, N = 253) = 289.25 became bigger, and all the fit indices were out of range.

The analyzed results of the proposed model by Socha and Sigler [13] were even worse. The latent variance-covariance matrix was not positively definite, which is a necessary condition for an acceptable model (see [28]). This was observed with the presence of the Heywood case in terms of a standardized correlation coefficient great than 1 between latent factors SI and ST. In a similar manner to M5, this model was modified, and the CFA results were reported with the heading Mod. M7 in Table 4. The resulting χ^2-value (89, N = 253) = 257.02, $p < 0.05$ is significant, but, when combined with GOF indices, qualifies the model to an appropriate fit of the data [29]. However, a comparison of this model results with the ones reported in Table 4 with the heading M9 [10] favored the latter. This is evident with the higher CFI/TLI and lower SRMR and RMSEA values observed in M9. Therefore, from the foregoing discussion, what appears to be the best explanation of the factor structure of the R-SPQ-2F in Norwegian context is the hypothesized model of López-Aguado and Gutiérrez-Provecho [10].

5. Conclusions

Teaching is considered successful when accompanied by meaningful learning. A good way to ensure that successful learning takes place is to investigate the approaches adopted by students when learning. Several efforts have been expended to promote deep learning in higher education such that the emerging leaders will better be prepared for an ever-increasingly diverse society [30,31]. Both qualitative and quantitative studies have, for several years, been directed towards proper conceptualization and operationalization of students' approaches to learning. Prominent among these studies are the works of Marton and Säljö [5,6], Entwistle and Waterston [32], and Biggs [33]. These have led to the development of measuring instruments in which the study process questionnaire (SPQ) seems to have gained global attention. However, studies on the cultural specificity of the latest SPQ called the R-SPQ-2F have generated diverse and inconclusive results.

In this article, investigations were geared towards addressing the issue of R-SPQ-2F cultural specificity when applied in the Norwegian context. Several models were compared, and what seems to be the best explanation of the R-SPQ-2F construct validity is a two first-order factor model involving deep and surface approaches to learning subscales. Meanwhile, by comparing the identified tested hypothesized model proposed in [10,13], as reported in Table 4, with the results of the appropriate fit (χ^2-value (151, N = 253) = 377.68, $p < 0.05$, SRMR = 0.072, CFI = 0.844, TLI= 0.824, and RMSEA = 0.077) found in [14], a conclusion could be drawn. There appears to be no obvious statistical difference with consideration for respective χ^2-values and GOF indices of these two models. Therefore, a two first-order factor model with 10 items measuring the deep approach and nine items (contrary to eight items in [10]) measuring the surface approach is still considered the best explanation for the R-SPQ-2F construct validity.

The justification for removing one item from the instrument was previously explained in detail in the first article. This study is to be followed up with an independent sample that will be collected in the near future for a confirmation of the proposed model and the predictive validity of the R-SPQ-2F. An interpretation of item scores can be achieved simply by adding corresponding items on the deep approach (scaled by dividing the sum by 10) and those on the surface approach (scaled by dividing the sum by 9) for classroom decisions. It is hoped that future replications of this study across other universities and groups of students will be carried out. This instrument is therefore recommended for measuring year-one undergraduate students' approaches in Norwegian universities. For instructional purposes, both data and Mplus syntax codes used for this study are available upon request from the author.

Funding: The article processing charge (APC) was funded by the University of Agder library.

Acknowledgments: The author acknowledge the support of the faculty of engineering and science, University of Agder, Kristiansand, Norway as well as MatRIC, centre for research, innovation, and coordination of mathematics teaching. Thanks to K. Bjørkestøl, H. K. Nilsen, and S. Goodchild for providing guidance and help with the Norwegian language as well as M. Lorås for providing some data.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Maciejewski, W.; Merchant, S. Mathematical tasks, study approaches, and course grades in undergraduate mathematics: A year-by-year analysis. *Int. J. Math. Educ. Sci. Technol.* **2016**, *47*, 373–387. [CrossRef]
2. Fryer, L.K.; Vermunt, J.D. Regulating approaches to learning: Testing learning strategy convergences across a year at university. *Br. J. Educ. Psychol.* **2018**, *88*, 21–41. [CrossRef] [PubMed]

3. Baeten, M.; Kyndt, E.; Struyven, K.; Dochy, F. Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educ. Res. Rev.* **2010**, *5*, 243–260. [CrossRef]
4. Biggs, J.B.; Tang, C. *Teaching for Quality Learning at University*, 3rd ed.; The Society for Research into Higher Education & Open University Press: Berkshire, UK, 2007.
5. Marton, F.; Säljö, R. On qualitative differences in learning II: Outcome as a function of the learner's conception of the task. *Br. J. Educ. Psychol.* **1976**, *46*, 115–127. [CrossRef]
6. Marton, F.; Säljö, R. On qualitative differences in learning I: Outcome and process. *Br. J. Educ. Psychol.* **1976**, *46*, 4–11. [CrossRef]
7. Biggs, J.B. *The Study Process Questionnaire (SPQ): Manual*; Australian Council for Educational Research: Hawthorn, Australia, 1987.
8. Biggs, J.B. What do inventories of students' learning processes really measure? A theoretical review and clarification. *Br. Educ. Res. J.* **1993**, *63*, 3–19. [CrossRef]
9. Biggs, J.B.; Kember, D.; Leung, D.Y.P. The revised two factor study process questionnaire: R-SPQ-2F. *Br. Educ. Res. J.* **2001**, *71*, 133–149. [CrossRef]
10. López-Aguado, M.; Gutiérrez-Provecho, L. Checking the underlying structure of R-SPQ-2F using covariance structure analysis/Comprobación de la estructura subyacente del R-SPQ-2F mediante análisis de estructura de covarianza. *Cult. Educ.* **2018**, *30*, 105–141. [CrossRef]
11. Chan, Y.M.; Sheung Chan, C.M. Approach to learning of sub-degree students in Hong Kong. *Res. Educ.* **2010**, *84*, 65–78. [CrossRef]
12. Stes, A.; De Maeyer, S.; Van Petegem, P. Examining the cross-cultural sensitivity of the revised two-factor study process questionnaire (R-SPQ-2F) and validation of a Dutch version. *PLoS ONE* **2013**, *8*, e54099. [CrossRef]
13. Socha, A.; Sigler, E.A. Exploring and "reconciling" the factor structure for the revised two-factor study process questionnaire. *Learn. Individ. Differ.* **2014**, *31*, 43–50. [CrossRef]
14. Zakariya, Y.F.; Bjørkestøl, K.; Nilsen, H.K.; Goodchild, S.; Lorås, M. University students' learning approaches: An. adaptation of the revised two-factor study process questionnaire to Norwegian. *Stud. Educ. Eval.* **2019**. under revision.
15. Justicia, F.; Pichardo, M.C.; Cano, F.; Berbén, A.B.; De La Fuente, J. The revised two-factor study process questionnaire (R-SPQ-2F): Exploratory and confirmatory factor analyses at item level. *Eur. J. Psychol. Educ.* **2008**, *23*, 355–372. [CrossRef]
16. Xie, Q. Validating the revised two-factor study process questionnaire among Chinese university students. *Int. J. Educ. Psychol. Assess.* **2014**, *16*, 4–21.
17. Merino, S.C.; Kumar, P.R. Validación estructural del R-SPQ-2F: Un análisis factorial confirmatorio. *Rev. Digit. Investig. Docencia Univ.* **2013**, *7*, 113–127. [CrossRef]
18. Önder, I.; Besoluk, S. Adaptation of revised two factor study process questionnaire (R-SPQ-2F) to Turkish. *Educ. Sci.* **2010**, *35*, 55–67.
19. Hu, L.T.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Modeling Multidiscip. J.* **1999**, *6*, 1–55. [CrossRef]
20. Hu, L.t.; Bentler, P.M. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol. Methods* **1998**, *3*, 424–453. [CrossRef]
21. Fryer, L.K.; Ginns, P.; Walker, R.A.; Nakao, K. The adaptation and validation of the CEQ and the R-SPQ-2F to the Japanese tertiary environment. *Br. J. Edu. Psychol.* **2012**, *82*, 549–563. [CrossRef] [PubMed]
22. Immekus, J.C.; Imbrie, P.K. A test and cross-validation of the revised two-factor study process questionnaire factor structure among western university students. *Educ. Psychol. Meas.* **2010**, *70*, 495–510. [CrossRef]
23. Suh, Y. The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Struct. Equ. Modeling Multidiscip. J.* **2015**, *22*, 568–580. [CrossRef]
24. Brown, T.A. *Confirmatory Factor Analysis for Applied Research*; The Guilford Press: New York, NY, USA, 2015.
25. Wolf, E.J.; Harrington, K.M.; Clark, S.L.; Miller, M.W. Sample size requirements for structural equation models: An. evaluation of power, bias, and solution propriety. *Educ. Psychol. Meas.* **2013**, *76*, 913–934. [CrossRef] [PubMed]
26. Muthén, L.K.; Muthén, B.O. *Mplus User's Guide*, 8th ed.; Muthén & Muthén: Los Angeles, CA, USA, 2017.

27. Bentler, P.M. Comparative fit indexes in structural models. *Psychol. Bull.* **1990**, *107*, 238–246. [CrossRef] [PubMed]
28. Byrne, B.M. *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*; Routledge, Taylor & Francis Group: New York, NY, USA, 2012.
29. Marsh, H.W.; Hau, K.T.; Wen, Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct. Equ. Modeling Multidiscip. J.* **2004**, *11*, 320–341. [CrossRef]
30. Lake, W.W.; Boyd, W.E. Is the university system in Australia producing deep thinkers? *Aust. Univ. Rev.* **2015**, *57*, 54–59.
31. Barattucci, M. Approach to study as an indicator of the quality of teaching and of learning environment: The contribution of John Biggs. *J. E Learn. Knowl. Soc.* **2017**, *13*, 77–88.
32. Entwistle, N.; Waterston, S. Approaches to studying and levels of processing in university students. *Br. Educ. Res. J.* **1988**, *58*, 258–265. [CrossRef]
33. Biggs, J.B. Individual and group differences in study processes. *Br. J. Educ. Psychol.* **1978**, *48*, 266–279. [CrossRef]

Paper III

Zakariya, Y. F., Goodchild, S., Bjørkestøl, K., & Nilsen, H. K. (2019). Calculus self-efficacy inventory: Its development and relationship with approaches to learning. *Education Sciences, 9*(3), 170. doi:10.3390/educsci9030170

education sciences

MDPI

Article

Calculus Self-Efficacy Inventory: Its Development and Relationship with Approaches to learning

Yusuf F. Zakariya *, Simon Goodchild, Kirsten Bjørkestøl and Hans K. Nilsen

Department of mathematical sciences, University of Agder, 4630 Kristiansand S, Norway
* Correspondence: yusuf.zakariya@uia.no

Received: 7 June 2019; Accepted: 25 June 2019; Published: 3 July 2019

Abstract: This study was framed within a quantitative research methodology to develop a concise measure of calculus self-efficacy with high psychometric properties. A survey research design was adopted in which 234 engineering and economics students rated their confidence in solving year-one calculus tasks on a 15-item inventory. The results of a series of exploratory factor analyses using minimum rank factor analysis for factor extraction, oblique promin rotation, and parallel analysis for retaining extracted factors revealed a one-factor solution of the model. The final 13-item inventory was unidimensional with all eigenvalues greater than 0.42, an average communality of 0.74, and a 62.55% variance of the items being accounted for by the latent factor, i.e., calculus self-efficacy. The inventory was found to be reliable with an ordinal coefficient alpha of 0.90. Using Spearman' rank coefficient, a significant positive correlation $\rho(95) = 0.27$, $p < 0.05$ (2-tailed) was found between the deep approach to learning and calculus self-efficacy, and a negative correlation $\rho(95) = -0.26$, $p < 0.05$ (2-tailed) was found between the surface approach to learning and calculus self-efficacy. These suggest that students who adopt the deep approach to learning are confident in dealing with calculus exam problems while those who adopt the surface approach to learning are less confident in solving calculus exam problems.

Keywords: self-efficacy; deep approach; surface approach; higher education; parallel analysis

1. Introduction

Studies on meaningful learning experiences of students in higher education have taken variant dimensions over the last decades. A good number of psychologists and sociologists have dug deep into students' reflections of themselves as they learn [1–3]. An outcome of this insight into students' learning is the identification of perceived self-efficacy as a good predictor of desirable learning outcomes [4]. Perceived self-efficacy, according to Bandura [5], refers to "beliefs in one's capabilities to organize and execute the courses of action required to manage prospective situations" (p. 2). These internal convictions put an individual in a better situation to approach a presented task and behave in a particular way. An individual will tend to engage in tasks for which they have perceived self-competence and try to avoid the ones with less perceived self-competence. Self-efficacy is a determinant factor that positively correlates with the amount of effort expended on a task, perseverance when faced with impediments, and resilience during challenging situations [1].

There has been a long-time debate among educationists on what are appropriate ways of assessing self-efficacy with some contending for the general perspective while others opting for the domain/situation specific perspective (e.g., [6,7]). The domain-specific perspective has influenced the conceptualization of self-efficacy around many fields. For example, mathematics self-efficacy has long been conceptualized as "a situational or problem-specific assessment of an individual's confidence in her or his fully perform or accomplish a particular" [2]. In a similar manner, engineering self-efficacy has been defined as a "person's belief that he or she can successfully navigate the engineering curriculum and eventually become a practicing engineer" [8]. Self-efficacy among engineering students has been investigated from conceptualization through developing measuring instruments to correlation with other variables like performance, anxiety, and performance [9,10]. In the same way it has been investigated in mathematics and other science-based courses.

Despite studies on mathematics self-efficacy and performance being sparse, especially in higher education (HE), the available empirical evidence has established a remarkable relationship between mathematics self-efficacy and academic performance, with the former being a strong predictor of the latter [11–14]. For example, Peters [15] reported a quantitative empirical study on the relationship between self-efficacy and mathematics achievement including other constructs among 326 undergraduate students. Employing multi-level analysis, it was found that mathematics self-efficacy differed across genders, with boys taking the lead, and positively correlated with achievement. More recently, Roick and Ringeisen [16] found, in their longitudinal study, that mathematics self-efficacy exerted a great influence on performance and played a mediating role between learning strategies and mathematics achievement. Similar corroborative results can also be found in the quantitative study reported in [17].

A good number of educators have empirically shown and emphatically argued that the best way to achieve a higher predictive power of mathematics self-efficacy on academic performance of students is through task-specific measures (e.g., [14]). Surprisingly, an extensive search of the literature revealed a lack of instruments for measuring students' self-efficacy on year-one calculus tasks. This is despite the fact that calculus has been a compulsory part of most year-one Science, Technology, Engineering, and Mathematics (STEM) curricula of many universities in the world. The current study therefore aimed at developing a measure for assessing students' self-efficacy on year-one calculus tasks with high psychometric properties. Furthermore, in order to enhance the predictive validity of the developed instrument its relationship with approaches to learning was also investigated.

2. Literature review

It is Albert Bandura who is considered the first psychologist in the history of clinical, social, and counseling psychology to have introduced the word "self-efficacy" (see, [18]) to refer to "the conviction that one can successfully execute the behavior required to produce the outcomes" [19]. However, some authors have contended that the "outcome expectancy" concept, which was extensively investigated prior to 1977, is equivalent to self-efficacy in theory, logic, and operationalization [20,21]. In Bandura's rebuttal of this criticism, he elicited the conceptual differences between outcome and self-efficacy expectancies while maintaining that the kinds of outcomes people expect are strongly influenced by self-efficacy expectancies (see, [22]). An overview of some of these controversies including arguments, counterarguments, disparities, and agreements can be found in the literature (e.g., [23,24]).

The basic tenet of the self-efficacy theory is that all psychological and behavioral changes occur as a result of modifications in the sense of efficacy or personal mastery of an individual [19,25]. In the words of Bandura [19], "people process, weigh, and integrate diverse sources of information concerning their capability, and they regulate their choice behavior and effort expenditure accordingly" (p. 212). In addition, Bandura's theory posits that the explanation and prediction of psychological changes can be achieved through appraisal of the self-efficacy expectations of an individual. In other words, the mastery or coping expectancy of an individual is a function of outcome expectancy—the credence that a given behavior will or will not result to a given outcome—and self-efficacy expectancy—"the belief that the person is or is not capable of performing the requisite" [23].

Furthermore, the applications of Bandura's theory as suitable frameworks of conceptualization are numerous in cardiac rehabilitation studies [26], educational research, clinical nursing, music and educational practices [27–30]. In a study involving undergraduate students taking a biomechanics course in the United States, Wallace and Kernozek [31] demonstrated how the self-efficacy theory can be used by instructors to improve students' learning experience and lower their anxiety towards the course. Moreover, Sheu et al. [32] reported a meta-analysis study on the contributions of self-efficacy theory in learning science, mathematics, engineering, and technology. The foregoing discussion points to the wide acceptance of Bandura's self-efficacy theory not only among psychologists but also the educational community at large.

The different conceptualizations of self-efficacy involving general and domain-specific perspectives have recurring implications on the measurement of the construct. A look into the literature reveals that mathematics self-efficacy has been measured with instruments tailored towards general assessment (e.g., [16]), sources of efficacy (e.g., [33]), task-specific efficacy (e.g., [34]), and adaptations from other instruments or which are self-developed (e.g., [35]). These instruments have their strengths and weaknesses. A brief account of each type of instruments is presented in the forthcoming paragraphs accompanied by the justification for a desired approach in the current study.

General assessment instruments have been developed to measure students' self-reported ratings of their capabilities to perform in mathematical situations. Chan and Yen Abdullah [36] developed a 14-item mathematics self-efficacy questionnaire (MSEQ) in which respondents appraised their ability on a five-point Likert scale from 1 (never) to 5 (usually). MSEQ had four sub-structures comprised of three items each measuring general mathematics self-efficacy and "efficacy in future" coupled with four items each measuring self-efficacy in class and in assignments. Evidence of validity was provided, and internal consistency of the items was investigated with Cronbach's alpha of 0.94, which showed high reliability. A similar result was also reported in an omnibus survey instrument developed by Wang and Lee [37], in which mathematics self-efficacy was a subcategory. These kinds of omnibus instruments have been reported to be problematic in their predictive relevance [38].

Other closely related instruments to mathematics general assessment types are the adapted mathematics subcategory items from other instruments. For example, in a longitudinal study involving 3014 students, You, Dang and Lim [39] developed a mathematics self-efficacy measure by adapting items from the motivated strategies for learning questionnaire (MSLQ) developed by Pintrich, Smith, Garcia, and McKeachie [40]. Furthermore, in an attempt to operationalize mathematics self-efficacy, Y.-L. Wang et al. [35] developed an instrument which was an adaptation of the science learning self-efficacy questionnaire developed in [41] by substituting mathematics for science in the original instrument. Some authors have independently developed measures for mathematics self-efficacy in which the sources of their items are not disclosed. For example, Skaalvik, Federici, and Klassen [42] developed a 4-item mathematics self-efficacy Norwegian measure as part of a survey instrument without any disclosure of the sources of their items. These instruments were not too different from the general academic self-efficacy measures in terms of their predictive power of performance [38].

Based on Bandura's [3,5] theorized sources of self-efficacy—*mastery experience*, *vicarious experience*, *verbal/social persuasions*, *physiological or affective states*—some educationists have developed and investigated some measures [33,43,44]. In a quantitative empirical three-phase study, Usher and Pajares [33] developed a measure and investigated the sources of mathematics self-efficacy. The study started in Phase One with an 84-item measure and ended in Phase Three with a revised 24-item instrument. The final version contained six items in each of the mastery experience, vicarious experience, social persuasions, and physiological state subcategories with 0.88, 0.84, 0.88, and 0.87 Cronbach's alpha coefficients as pieces of evidence of item internal consistency, respectively. The study confirmed the hypothesized mastery experience of Bandura [5] as the strongest predictor of learning outcome [33]. Other studies have also reported corroborative empirical evidence to confirm the hypothesized sources of mathematics self-efficacy using Usher and Pajares' [33] instrument with either wording or language adaptations [45,46].

With the exception of sources of self-efficacy measures, the most effective approach in terms of achieving high predictive power of learning outcome is to assess mathematics self-efficacy through a task-specific measure [47]. The basic idea in developing a mathematics task-specific instrument is to conceptualize self-efficacy on predefined mathematical task(s) and tailor the instrument items towards the respondent's self-capability to complete the tasks. An example of early instruments developed using this approach was the 52-item mathematics self-efficacy scale (MSES) by Betz and Hackett [34] to measure self-efficacy among mathematics college students. In the administration of this instrument, the respondents had to rate their confidence in successfully completing 18-item mathematics tasks; solving 18-item math related problems; and achieving at least a "B" grade in a 16-item college mathematics related course like calculus, statistics, etc. Evidence of reliability was provided with Cronbach's alpha coefficients of 0.90, 0.93, and 0.92 on each subscale as well as 0.96 on the 52-item scale [34]. MSES has been investigated, revised, and validated with items adapted to university mathematics tasks/problems as well as its rating reduced from a 10-point to five-point Likert scale [14,48].

A task-specific mathematics self-efficacy instrument was also utilized by the Programme for International Student Assessment (PISA) in their 2012 international survey across 65 countries as reported in [49]. The eight-item instrument measured students' self-reported level of confidence in completing some mathematical tasks without solving the problems. The rating involved a five-point Likert scale ranging from "not at all confident" to "very confident"' in which students were asked, for example, "how confident would they feel about solving an equation like $2(x + 3) = (x + 3)(x - 3)$"? Cronbach's alpha coefficient of 0.83 was provided as evidence of reliability [49].

3. Methods

3.1. Item Development

The items of the calculus self-efficacy inventory (CSEI) were developed based on the recommendations of Bandura's self-efficacy theory using the guidelines explained in the literature (e.g., [50]). The initial inventory used in the current study contained 15 items selected from old final examination questions in a year-one calculus course from 2014/2015 to 2018/2019 academic sessions. Some of the topics covered in the course were functions, limits, continuity and differentiability, differentiations and its applications, integration and its applications, etc. The items varied in level of difficulty from procedural (involving recall of facts, definition, use of formulae, etc.) to conceptual items which involve higher cognitive abilities such as applications, analysis, evaluations, etc. The students were asked to rate their confidence to solve the tasks on a scale ranging from 0 (not confident at all), through 50 (moderate confidence), to 100 (very confident). The 100-point scale was used because it has been reported to enhance the predictive validity of the self-efficacy inventory (see, [50]). Sample questions are presented in Table 1.

Table 1. Sample items on the calculus self-efficacy inventory (CSEI).

SN.	How Confident are You that You can Solve Each of These Problems Right Now?	Confidence (0–100)
3	Calculate the limit: $\lim_{x\to 1} \frac{1-\cos(1-x^2)}{x^2-2x+1}$	
7	A curve is given by $x = y^2 - x^2y - 1$. Use implicit differentiation to find y'.	
11	Evaluate the integral. $\int \frac{x-7}{x^2+x-6}dx$	
14	A surface is bounded by the function $f(x) = \frac{1}{3}e^{x^2}$ where $0 \leq x \leq 2$ and the *x-axis*. A vessel is made by rotating the surface about the *x-axis*. Find the resulting volume.	

3.2. Research Design and Participants

This study adopted a survey research design involving 234 year-one university students in engineering and economics programs offering a compulsory calculus course. The study population comprised 135 males and 99 females with an average age between 19–22 years. The multicollinearity and adequacy of the sample correlation matrix was checked using Bartlett's test sphericity (N = 234, d = 91) = 1632.2, $p < 0.05$, which was significant, with a Kaiser–Meyer–Olkin (KMO) test = 0.88 and a determinant greater than 0.00001. These all confirmed the sufficiency of the sample for factor analysis as well as absence of multicollinearity in the data [51]. Moreover, the sample was also within the suggested ranges, in the literature, for factor analysis of multiple item instruments (e.g., [52]).

3.3. Materials

Two instruments were used in this study. The first was the 15-item CSEI described in the previous section entitled "item development". The second instrument was a Norwegian version of the two-factor revised study process questionnaire (R-SPQ-2F) developed by Biggs, Kember, and Leung [53]. This version is a 19-item instrument that measures students' approaches to learning on a five-point Likert scale with 10 items measuring deep approach to learning and 9 items measuring surface approach to learning mathematics. The psychometric properties of this instrument were investigated elsewehere [54,55], and its reliability was found to be appropriate from 0.72 to 0.81 using Raykov and Marcoulides' [56] formula.

3.4. Procedure

The data were collected using both electronic and paper versions of the two questionnaires. A total of 110 engineering students completed both the CSEI and the R-SPQ-2F, out of whom 95 gave us their consent to identify their scores on both scales. Economics students only completed the CSEI due to some logistic problems and formed the remaining 124 of the sample. The collected data were screened for outlier cases and found to contain none. Responses on CSEI were coded on a 11-point scale with 0 coded as 0, $0 <$ values ≤ 10 coded as 1, $10 <$ values ≤ 20 coded as 2, ... , and $90 <$ values ≤ 100 coded as 11. Univariate and multivariate descriptive statistics analysis of the data revealed the presence of excess kurtosis and skewness as both indices were greater than $|1.0|$ on most of the items of CSEI [57]. For this reason, 11-point categories were further collapsed to five-point ones, and a polychoric correlation matrix was used in the factor analysis of the data using the FACTOR program version 10.8.04 [58]. The recoding into five-point categories was done is such a way that 0–2 were coded as 1, 3–4 were coded as 2, ... , and 9–10 were coded as 5.

4. Results

4.1. Factor Analysis of CSEI

An exploratory factor analysis (EFA) was run on the 15-item CSEI data to determine the factor structures of the inventory. As the data were found to contain excess kurtosis and skewness, instead of a Pearson correlation matrix, a polychoric correlation matrix was used to enhance analysis effectiveness [59]. Minimum rank factor analysis (MRFA) was used in extracting the common underlying factors of CSEI instead of maximum likelihood (ML), unweighted least squares, etc., due to its ability to optimally yield communalities of the sample covariance matrix [60]. The number of factors to retain was based on the optimized parallel analysis procedure [61,62] which has been confirmed to outperform the original Horn's parallel analysis [63].

This procedure involves simulations of 500 datasets by permuting the sample data at random so that numbers of cases and variables are unchanged. On each of these datasets, EFA was conducted using MRFA, and the average eigenvalues of the extracted factors were then compared with the eigenvalues of the sample. Factors with eigenvalues greater than the average eigenvalues of the simulated datasets were then retained. This procedure has been shown to be an effective way of deciding the number of factors to retain in EFA and also outperformed Kaiser's criteria of eigenvalues greater than 1 and use of scree plot [61]. The extracted factors were rotated using promin, an example of oblique rotations described in [64]. An oblique rotation was appropriate because the latent factors are assumed to be correlated contrary to the assumption of disjoint factors in the orthogonal rotations. The analysis was performed on both the 11-point and five-point coding of the data. However, results from the five-point coding are presented in Table 2 due to slightly higher precisions in estimating factor loadings and communalities of the items. Factor loadings less than or equal to |0.30| are excluded from Table 2.

Table 2. Rotated and unrotated factor loadings and item communalities.

CSEI	First Analysis		Second Analysis		Third Analysis	
	F1	F2	F1	Communality	F1	Communality
Item 01		0.99	0.73	1.00	——	——
Item 02		0.51	0.50	0.82	0.43	0.42
Item 03		0.61	0.78	0.72	0.78	0.83
Item 04	0.79		0.62	0.77	0.64	0.70
Item 05	0.34		0.51	0.48	0.51	0.50
Item 06	0.83		0.55	0.80	0.60	0.89
Item 07	0.40	0.35	0.66	0.60	0.65	0.55
Item 08		0.71	0.76	0.82	0.73	0.75
Item 09	0.85	−1.04	——	——-	——	——
Item 10	0.35	0.38	0.65	0.75	0.65	0.74
Item 11	0.38		0.74	0.95	0.76	0.96
Item 12	0.69		0.84	0.95	0.85	0.91
Item 13	1.02	−0.30	0.68	0.98	0.72	0.90
Item 14		0.80	0.74	0.77	0.72	0.77
Item 15	0.45		0.68	0.77	0.68	0.66

Table 2 presents rotated and unrotated factor loadings of a series of three exploratory factor analyses of the CSEI data. The first analysis column of Table 2 represents rotated factor loadings of a two-factor solution of the data. However, there was a gross misspecification in this model with Items 07, 09, and 13 exhibiting substantial cross-loadings and out of range rotated factor loadings. The out of range factor loadings in Item 09 (−1.04) and Item 13 (1.02) are suggestive of negative error variance in the factor solutions of the items. Furthermore, a look at the polychoric correlation matrix (see Appendix A) also revealed that Item 09 had negative correlation coefficients with most other items, which is an indication of a negative variance. For this reason, Item 09 was deleted before the second EFA was run. Moreover, the results of the optimized parallel analysis (Table 3) recommended retaining one-factor solution in the model based on the 95 percentile and 2-factor solution based on the mean. However, the 95 percentile recommendation of the parallel analysis has been reported to be more accurate than its recommendation based on the mean [61]. Therefore, the second analysis was run with a fixed one-factor solution of the model.

Table 3. Parallel analysis—minimum rank factor analysis (MRFA) results based on the polychoric correlation matrix.

Variable	Real-Data % of Variance	Mean of Random % of Variance	95 Percentile Random % of Variance
1	50.09 **	17.00	19.33
2	17.04 *	15.25	17.18
3	6.46	13.75	15.24
4	5.47	12.15	13.33
5	5.27	10.51	11.84
6	4.25	8.81	10.19
7	4.18	7.37	8.79
8	2.92	5.99	7.26
9	2.17	4.60	5.83
10	1.23	3.24	4.38

** Advised number of dimensions when 95 percentile is considered: 1. * Advised number of dimensions when mean is considered: 2.

The second analysis column of Table 2 presents unrotated factor loadings and item communalities of a one-factor solution of the model with the exclusion of Item 09. This solution contains a Heywood case in form of the communality of Item 01 equals 1. This means that all the variance of Item 01 is shared with other items in the model and that this item has no unique variance at all [51]. Item 01 was removed from the model for this reason and the third analysis was run. The third analysis column of Table 2 presents unrotated factor loadings and item communalities of a one-factor solution of the model excluding Items 01 and 09. All factor loadings were greater than 0.42 and the average communality (0.74) was greater than the widely recommended 0.70, which are suggestive of a good model solution for the sample data [65]. The extracted eigenvalues accounted for a total of 62.55% common variance as depicted in Table 4. This can be interpreted to mean that the one-factor model explained 62.55% of common variance of the factor solution which can be used to justify goodness of fit of the model.

Table 4. Eigenvalues and proportion of explained variance.

Variable	Eigenvalue	Proportion of Common Variance	Cumulative Proportion of Variance	Cumulative Percentage of Variance
1	5.9848	0.6255	0.6255	62.55
2	1.2650	0.1322		
3	0.7748	0.0810		
4	0.4287	0.0448		
5	0.3477	0.0363		
6	0.3258	0.0341		
7	0.1981	0.0207		
8	0.1438	0.0150		
9	0.0911	0.0095		
10	0.0087	0.0009		
11	0.0001	0.0000		
12	0.0000	0.0000		
13	0.0000	0.0000		

4.2. Reliability of the Instrument

There have been heated debates among methodologists on the appropriateness of using Cronbach's alpha coefficients in estimating reliability of ordinal scale data. Some of these debates have been provoked by gross misuses and misinterpretations of Cronbach's alpha especially in the presence of excess kurtosis and skewness, violations of the normality assumption, non-continuous item level of measurement, etc., inherent in ordinal data [66,67]. To circumvent this problem, alternative indices have been proposed for estimating the reliability of ordinal scales (e.g., [68,69]).

A widely used alternative estimate of reliability is the ordinal coefficient alpha proposed by Zumbo, Gadermann, and Zeisser [70]. Ordinal coefficient alpha is similar to Cronbach's alpha coefficient in that they are both computed using the McDonald's [71] formula (Equation (1)) for a one-factor factor analysis model. However, the former is based on polychoric correlation matrix estimates that are theoretically different from the Pearson correlation matrix estimates used in the latter. It has been shown both through simulation and raw data studies that ordinal coefficient alpha outperforms the Cronbach's alpha coefficient in estimating reliability of scales measured using the Likert format of fewer than six-point categories (e.g., [70,72]).

$$\alpha = \frac{p}{p-1}\left[\frac{p * \overline{\lambda}^2 - \overline{c}}{p * \overline{\lambda}^2 + \overline{u}}\right] \quad (1)$$

In Equation (1), α is the ordinal coefficient, p is the number of items in the instrument, and $\overline{\lambda}$, $\overline{c}$ and $\overline{u}$ (where $u = 1 - c$) are the average factor loading, average communality, and average unique variance, respectively. Using the values of these parameters as presented in Table 5, the ordinal coefficient alpha can be calculated as follows:

$$\alpha = \frac{13}{13-1}\left[\frac{13 * .67^2 - .74}{13 * .67^2 + 0\ .26}\right] = 0.91. \tag{2}$$

Table 5. Ordinal coefficient alpha reliability parameters.

CSEI	λ	c	u
Item 02	0.43	0.42	0.58
Item 03	0.78	0.83	0.17
Item 04	0.64	0.70	0.30
Item 05	0.51	0.50	0.50
Item 06	0.60	0.89	0.11
Item 07	0.65	0.55	0.45
Item 08	0.73	0.75	0.25
Item 10	0.65	0.74	0.26
Item 11	0.76	0.96	0.04
Item 12	0.85	0.91	0.09
Item 13	0.72	0.90	0.10
Item 14	0.72	0.77	0.23
Item 15	0.68	0.66	0.34
Average	**0.67**	**0.74**	**0.26**

This is suggestive of a highly reliable unidimensional instrument with an appropriate internal item consistency.

4.3. Correlation of Calculus Self-Efficacy with Approaches to Learning

In an effort to examine the predictive validity of the CSEI, a correlation between students' scores on the inventory and their respective scores on the R-SPQ-2F was investigated. Scoring of the CSEI was accomplished by adding item scores on the final 13-item inventory while that of R-SPQ-2F was in line with the procedure described in [54]. Each of the 95 engineering students had scores on self-efficacy and deep and surface approaches to learning. These scores were explored using descriptive statistics and tested for normality assumptions before the correlation analysis. As shown in Table 6 and Figure 1, scores on both deep and surface approaches are normally distributed while scores on CSEI are not.

Table 6. Descriptive statistics and Shapiro–Wilk's test of normality results.

	Descriptive Statistics									Test of Normality		
	N	Min.	Max.	Mean	Std. Dev.	Skewness		Kurtosis		Shapiro–Wilk		
						Stat.	Std. Err.	Stat.	Std. Err.	Stat.	df	Sig.
Deep approach	95	1.20	4.70	2.82	0.68	0.03	0.25	0.10	0.49	0.99	95	0.82
Surface approach	95	1.00	4.00	2.42	0.63	0.15	0.25	-0.49	0.49	0.99	95	0.65
CSEI	95	13.00	65.00	46.07	11.43	−0.92	−0.25	0.86	0.49	0.94	95	0.00*

* Significant, $p < 0.05$.

The non-normal distribution of scores in the CSEI is evident from the significance level of Shapiro–Wilk's test statistic (N = 95, df = 95) = 0.94, p < 0.05, as shown in Table 6. Furthermore, the CSEI scores also exhibited a negatively skewed distribution as shown in the last diagram of Figure 1. For these reasons, a nonparametric bivariate Spearman rank correlation was used instead of the Pearson correlation to check the relationship between the CSEI and the R-SPQ-2F scores. The results revealed a significant positive correlation $\rho(95) = 0.27$, $p < 0.05$ (2-tailed) between the deep approach to learning and calculus self-efficacy and a significant negative correlation $\rho(95) = -0.26$, $p < 0.05$ (2-tailed) between the surface approach to learning and calculus self-efficacy. These results could be interpreted to mean, at the group level, that students who adopt the deep approach to learning are usually confident in dealing with calculus exam problems while those who adopt the surface approach to learning are less confident to successfully solve calculus exam problems. This finding confirms the hypothesis of the Bandura's self-efficacy theory [4,6] and also corroborates the mediating role played by self-efficacy between learning strategies and performance reported in [16].

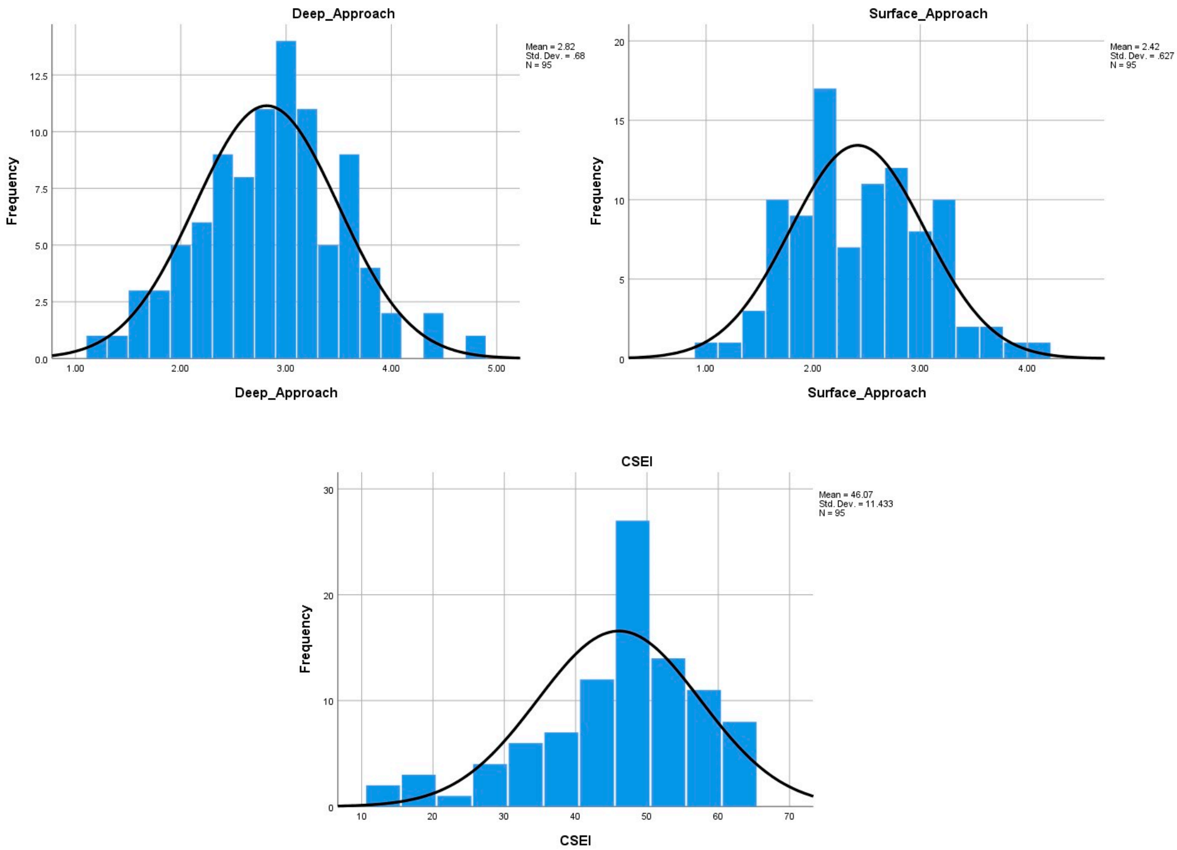


Figure 1. Normal distribution of scores on CSEI and R-SPQ-2F scales.

5. Conclusions

Despite the abundant empirical evidence of the high predictive power of task-specific mathematics self-efficacy in the literature, an instrument for its measure is still lacking [4,50]. The current study was framed within a quantitative research methodology to develop a concise measure of calculus self-efficacy with high psychometric properties among year-one university students. Bandura's self-efficacy theory provided a theoretical framework for the conceptualization and operationalization of items on the developed calculus self-efficacy inventory (CSEI). This theory posits that all psychological and behavioral changes occur as a result of modifications in the sense of efficacy or personal mastery of an individual [19,25]. On this basis, the accompanied guidelines and recommendations of this theory [50] were followed in constructing the CSEI items.

The initial instrument contained 15 items, in which 234 respondents rated their confidence in solving year-one calculus tasks on a 100-point rating scale. The results of the factor analysis using MRFA for factor extraction, promin rotation, and parallel analysis for retaining factors revealed a one-factor solution of the model. The final 13-item inventory was unidimensional with all eigenvalues greater than 0.42, an average communality of 0.74, and a 62.55% variance of the items being accounted for by the latent factor, i.e., calculus self-efficacy. These results can be interpreted as evidence of construct validity in measuring students' internal confidence in successfully solving some calculus tasks. The CSEI has the following advantages over the mathematics self-efficacy scale (MSES) developed by Betz and Hackett [34] and its revisions (e.g., [48]): Its concise length, task specificity, higher factor loadings, and communality.

Furthermore, the reliability coefficient of the CSEI was found to be 0.91 using the ordinal coefficient alpha with the formula described in [70]. This coefficient portrays evidence of high internal consistency of items in the inventory [63]. This reliability coefficient is higher than the coefficient of the mathematics task subscale of the MSES reported in [2,34], and it is within the ranges of the revised MSES reported in [14,48]. There are some misconceptions on the appropriate use of the ordinal coefficient alpha for estimating scale reliability as can be found in [73]. These misconceptions are acknowledged. However, the examples of the types of items provided in Chalmer's own article are enough to justify the use of the ordinal coefficient alpha in the current study.

The results of the current study also provided an insight into the correlation between approaches to learning and calculus self-efficacy. The significant positive correlation between the deep approach and self-efficacy as well the significant negative correlation between the surface approach and self-efficacy are indications of the predictive validity of the CSEI. This finding also confirms the hypothesis of Bandura's self-efficacy theory [4,6] as well as corroborates the mediating role played by self-efficacy between learning strategies and performance reported in [16]. It is a crucial to remark that the causal effect between calculus self-efficacy and approaches to learning is not claimed with this finding. Rather, the results have only established a relationship between these constructs that can be explored further in future studies. The final 13-item instrument is available in English and Norwegian upon request from the corresponding author. This inventory is therefore recommended to university teachers in order to assess students' confidence in successfully solving calculus tasks.

Author Contributions: Y.F.Z. made contributions in Conceptualization, Data curation, Formal analysis, Methodology, Writing-original draft, and writing-review & editing. S.G. made contributions in Methodology, Resources, Supervision, Writing-review & editing. K.B. made contributions in Methodology, Resources, Supervision, Writing-review & editing. H.K.N. made contributions in Methodology, Resources, Supervision, Writing-review & editing.

Funding: The article processing charge (APC) was funded by the University of Agder library.

Acknowledgments: The Faculty of Engineering and Science, University of Agder, Kristiansand, Norway as well as the MatRIC, Centre for Research, Innovation and Coordination of Mathematics Teaching are acknowledged for supporting this research.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Standardized Variance/Covariance Matrix (Polychoric Correlation)

Variable	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
CSEI 01	1.000														
CSEI 02	0.642	1.000													
CSEI 03	0.682	0.427	1.000												
CSEI 04	0.255	0.227	0.501	1.000											
CSEI 05	0.365	0.210	0.409	0.368	1.000										
CSEI 06	0.099	0.152	0.377	0.661	0.232	1.000									
CSEI 07	0.502	0.314	0.543	0.339	0.266	0.367	1.000								
CSEI 08	0.665	0.307	0.602	0.352	0.409	0.356	0.520	1.000							
CSEI 09	-0.527	-0.136	-0.302	0.220	0.054	0.241	-0.064	-0.341	1.000						
CSEI 10	0.463	0.328	0.499	0.366	0.435	0.323	0.383	0.612	-0.026	1.000					
CSEI 11	0.373	0.253	0.477	0.492	0.266	0.440	0.517	0.376	0.200	0.398	1.000				
CSEI 12	0.547	0.301	0.632	0.446	0.305	0.459	0.543	0.573	-0.014	0.459	0.756	1.000			
CSEI 13	0.201	0.220	0.391	0.506	0.340	0.501	0.401	0.286	0.327	0.333	0.826	0.709	1.000		
CSEI 14	0.684	0.346	0.623	0.280	0.444	0.153	0.502	0.667	-0.405	0.480	0.400	0.609	0.355	1.000	
CSEI 15	0.393	0.331	0.502	0.422	0.365	0.387	0.433	0.536	0.030	0.361	0.375	0.555	0.395	0.601	1.000

References

1. Pajares, F. Self-efficacy beliefs in academic settings. *Rev. Educ. Res.* **1996**, *66*, 543–578. [CrossRef]
2. Hackett, G.; Betz, N.E. An exploration of the mathematics self-efficacy/mathematics performance. *J. Res. Math. Educ.* **1989**, *20*, 261–273. [CrossRef]
3. Bandura, A. The explanatory and predictive scope of self-efficacy theory. *J. Soc. Clin. Psychol.* **1986**, *4*, 359–373. [CrossRef]
4. Bandura, A. Perceived self-efficacy in cognitive development and functioning. *Educ. Psychol.* **1993**, *28*, 117–148. [CrossRef]
5. Bandura, A. *Self-efficacy: The Exercise of Control*; W H Freeman: New York, NY, USA, 1997.
6. Bandura, A. On the functional properties of perceived self-efficacy revisited. *J. Manag.* **2012**, *38*, 9–44. [CrossRef]
7. Scherbaum, C.A.; Cohen-Charash, Y.; Kern, M.J. Measuring general self-efficacy: A comparison of three measures using item response theory. *Educ. Psychol. Meas.* **2006**, *66*, 1047–1063. [CrossRef]
8. Jordan, K.L.; Sorby, S.; Amato-Henderson, S.; Donahue, T.H. Engineering Self-Efficacy of Women Engineering Students at Urban vs. Rural Universities. In Proceedings of the 41st ASEE/IEEE Frontiers in Education Conference, Rapid City, SD, USA, 12–15 October 2010.
9. Marra, R.M.; Schuurman, M.; Moore, C.; Bogue, B. Women Engineering Students' Self-Efficacy Beliefs—The Longitudinal Picture. In Proceedings of the American Society for Engineering Education Annual Conference and Exposition, Portland, OR, USA, 12–15 June 2005.
10. Carberry, A.R.; Lee, A.-S.; Ohland, M.W. Measuring engineering design self-efficacy. *J. Eng. Educ.* **2010**, *99*, 71–79. [CrossRef]
11. Jaafar, W.M.W.; Ayub, A.F.M. Mathematics self-efficacy and meta-cognition among university students. *Procedia-Soc. Behav. Sci.* **2010**, *8*, 519–524. [CrossRef]
12. Yusuf, M. The impact of self-efficacy, achievement motivation, and self-regulated learning strategies on students' academic achievement. *Procedia-Soc. Behav. Sci.* **2011**, *15*, 2623–2626. [CrossRef]
13. Akin, A.; Kurbanoglu, I.N. The Relationships between math anxiety, math attitudes, and self-efficacy: A structural equation model. *Studia Psychol.* **2011**, *53*, 263–273.
14. Pajares, F.; Miller, M.J. Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *J. Couns. Psychol.* **1995**, *42*, 190–198. [CrossRef]
15. Peters, M.L. Examining the relationships among classroom climate, self-efficacy, and achievement in undergraduate mathematics: A multi-level analysis. *Int. J. Sci. Math. Educ.* **2013**, *11*, 459–480. [CrossRef]
16. Roick, J.; Ringeisen, T. Students' math performance in higher education: Examining the role of self-regulated learning and self-efficacy. *Learn. Individ. Differ.* **2018**, *65*, 148–158. [CrossRef]
17. Lin, L.; Lee, T.; Snyder, L.A. Math self-efficacy and STEM intentions: A person-centered approach. *Front. Psychol.* **2018**, *9*, 2033. [CrossRef] [PubMed]
18. Kirsch, I. Early research on self-Efficacy: What we already know without knowing we knew. *J. Soc. Clin. Psychol.* **1986**, *4*, 339–358. [CrossRef]
19. Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* **1977**, *84*, 191–215. [CrossRef]
20. Kirsch, I. Self-efficacy and expectancy: Old wine with new labels. *J. Personal. Soc. Psychol.* **1985**, *49*, 824–830. [CrossRef]
21. Eastman, C.; Marzillier, J.S. Theoretical and methodological difficulties in Bandura's self-efficacy theory. *Cogn. Ther. Res.* **1984**, *8*, 213–229. [CrossRef]
22. Bandura, A. Recycling misconceptions of perceived self-efficacy. *Cogn. Ther. Res.* **1984**, *8*, 231–255. [CrossRef]
23. Maddux, J.E.; Sherer, M.; Rogers, R.W. Self-efficacy expectancy and outcome expectancy: Their relationship and their effects on behavioral intentions. *Cogn. Ther. Res.* **1982**, *6*, 207–211. [CrossRef]
24. Maddux, J.E.; Stanley, M.A. Self-efficacy theory in contemporary psychology: An overview. *J. Soc. Clin. Psychol.* **1986**, *4*, 244–255. [CrossRef]
25. Bandura, A. Self-efficacy mechanism in human agency. *Am. Pyschol.* **1982**, *37*, 122–147. [CrossRef]
26. Jeng, C.; Braun, L.T. Bandura's self-efficacy theory: A guide for cardiac rehabilitation nursing practice. *J. Holist. Nurs.* **1994**, *12*, 425–436. [CrossRef] [PubMed]

27. Montcalm, D.M. Applying Bandura's theory of self-efficacy to the teaching of research. *J. Teach. Soc. Work* **1999**, *19*, 93–107. [CrossRef]
28. Artino, A.R. Academic self-efficacy: From educational theory to instructional practice. *Perspect. Med. Educ.* **2012**, *1*, 76–85. [CrossRef] [PubMed]
29. Kardong-Edgren, S. Bandura's self-efficacy theory . . . Something is missing. *Clin. Simul. Nurs.* **2013**, *9*, e327–e328. [CrossRef]
30. Hendricks, K.S. The sources of self-efficacy: Educational research and implications for music. *Update Appl. Res. Music Educ.* **2016**, *35*, 32–38. [CrossRef]
31. Wallace, B.; Kernozek, T. Self-efficacy theory applied to undergraduate biomechanics instruction. *J. Hosp. Leis. Sport Tour. Educ.* **2017**, *20*, 10–15. [CrossRef]
32. Sheu, H.-B.; Lent, R.W.; Miller, M.J.; Penn, L.T.; Cusick, M.E.; Truong, N.N. Sources of self-efficacy and outcome expectations in science, technology, engineering, and mathematics domains: A meta-analysis. *J. Vocat. Behav.* **2018**, *109*, 118–136. [CrossRef]
33. Usher, E.L.; Pajares, F. Sources of self-efficacy in mathematics: A validation study. *Contemp. Educ. Psychol.* **2009**, *34*, 89–101. [CrossRef]
34. Betz, N.E.; Hackett, G. The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *J. Vocat. Behav.* **1983**, *23*, 329–345. [CrossRef]
35. Wang, Y.-L.; Liang, J.-C.; Lin, C.-Y.; Tsai, C.-C. Identifying Taiwanese junior-high school students' mathematics learning profiles and their roles in mathematics learning self-efficacy and academic performance. *Learn. Individ. Differ.* **2017**, *54*, 92–101. [CrossRef]
36. Chan, H.Z.; Yen Abdullah, M.N.L. Validity and reliability of the mathematics self-efficacy questionnaire (MSEQ) on primary school students. *Pertanika J. Soc. Sci. Humanit.* **2018**, *26*, 2161–2177.
37. Wang, X.; Lee, Y.S. Investigating the psychometric properties of a new survey instrument measuring factors related to upward transfer in STEM fields. *Rev. High. Educ.* **2019**, *42*, 339–384. [CrossRef]
38. Mamaril, N.A.; Usher, E.L.; Li, C.R.; Economy, D.R.; Kennedy, M.S. Measuring undergraduate students' engineering self-efficacy: A validation study. *J. Eng. Educ.* **2016**, *105*, 366–395. [CrossRef]
39. You, S.; Dang, M.; Lim, S.A. Effects of student perceptions of teachers' motivational behavior on reading, english, and mathematics achievement: The mediating role of domain specific self-efficacy and intrinsic motivation. *Child Youth Care Forum* **2016**, *45*, 221–240. [CrossRef]
40. Pintrich, P.R.; Smith, D.A.F.; Garcia, T.; McKeachie, W.J. Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educ. Psychol. Meas.* **1993**, *53*, 801–813. [CrossRef]
41. Lin, T.-J.; Tsai, C.-C. A multi-dimensional instrument for evaluating Taiwanese high school students' science learning self-efficacy in relation to their approaches to learning science. *Int. J. Sci. Math. Educ.* **2013**, *11*, 1275–1301. [CrossRef]
42. Skaalvik, E.M.; Federici, R.A.; Klassen, R.M. Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *Int. J. Educ. Res.* **2015**, *72*, 129–136. [CrossRef]
43. Lent, R.W.; Lopez, F.G.; Bieschke, K.J. Mathematics self-efficacy: Sources and relation to science-based career choice. *J. Couns. Psychol.* **1991**, *38*, 424–430. [CrossRef]
44. Joët, G.; Usher, E.L.; Bressoux, P. Sources of self-efficacy: An investigation of elementary school students in France. *J. Educ. Psychol.* **2011**, *103*, 649–663. [CrossRef]
45. Yurt, E. The predictive power of self-efficacy sources for mathematics achievement. *Educ. Sci.* **2014**, *39*, 176. [CrossRef]
46. Zientek, L.R.; Fong, C.J.; Phelps, J.M. Sources of self-efficacy of community college students enrolled in developmental mathematics. *J. Furth. High. Educ.* **2019**, *43*, 183–200. [CrossRef]
47. Toland, M.D.; Usher, E.L. Assessing mathematics self-efficacy. *J. Early Adolesc.* **2015**, *36*, 932–960. [CrossRef]
48. Kranzler, J.H.; Pajares, F. An exploratory factor analysis of the mathematics self-efficacy scale revised (MSES-R). *Meas. Eval. Couns. Dev.* **1997**, *29*, 215–228.
49. Borgonovi, F.; Pokropek, A. Seeing is believing: Task-exposure specificity and the development of mathematics self-efficacy evaluations. *J. Educ. Psychol.* **2018**, *111*, 268–283. [CrossRef]
50. Bandura, A. Guide for constructing self-efficacy scales. In *Self-Efficacy Beliefs of Adolescents*; Pajares, F., Urdan, T., Eds.; Information Age Publishing: Greenwich, CT, USA, 2006; Volume 5, pp. 307–337.
51. Field, A. *Discovering Statistics Using IBM SPSS*, 5th ed.; SAGE Publication Ltd.: London, UK, 2018.

52. Gagne, P.; Hancock, G.R. Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivar. Behav. Res.* **2006**, *41*, 65–83. [CrossRef]
53. Biggs, J.B.; Kember, D.; Leung, D.Y.P. The revised two factor study process questionnaire: R-SPQ-2F. *Br. Educ. Res. J.* **2001**, *71*, 133–149. [CrossRef]
54. Zakariya, Y.F.; Bjørkestøl, K.; Nilsen, H.K.; Goodchild, S.; Lorås, M. University students' learning approaches: An adaptation of the revised two-factor study process questionnaire to Norwegian. *Stud. Educ. Eval.* **2019**. under review.
55. Zakariya, Y.F. Study approaches in higher education mathematics: Investigating the statistical behaviour of an instrument translated into Norwegian. *Educ. Sci.* **2019**. under review.
56. Raykov, T.; Marcoulides, G.A. Scale reliability evaluation under multiple assumption violations. *Struct. Equ. Model. A Multidiscip. J.* **2016**, *23*, 302–313. [CrossRef]
57. Muthén, B.O.; Kaplan, D. A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *Br. J. Math. Stat. Psychol.* **1992**, *45*, 19–30. [CrossRef]
58. Lorenzo-Seva, U.; Ferrando, P.J. FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Appl. Psychol. Meas.* **2013**, *37*, 497–498. [CrossRef]
59. Holgado-Tello, F.P.; Chacón-Moscoso, S.; Barbero-García, I.; Vila-Abad, E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual. Quant.* **2008**, *44*, 153–166. [CrossRef]
60. Shapiro, A.; ten Berge, J.M.F. Statistical inference of minimum rank factor analysis. *Psychometrika* **2002**, *67*, 79–94. [CrossRef]
61. Timmerman, M.E.; Lorenzo-Seva, U. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol. Methods* **2011**, *16*, 209–220. [CrossRef] [PubMed]
62. Buja, A.; Eyuboglu, N. Remarks on Parallel Analysis. *Multivar. Behav. Res.* **1992**, *27*, 509–540. [CrossRef] [PubMed]
63. Baglin, J. Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Pract. Assess. Res. Eval.* **2014**, *19*, 1–15.
64. Lorenzo-Seva, U. Promin: A Method for Oblique Factor Rotation. *Multivar. Behav. Res.* **1999**, *34*, 347–365. [CrossRef]
65. Pituch, A.K.; Stevens, J.P. *Applied Multivariate Statistics for the Social Sciences*, 6th ed.; Routledge: New York, NY, USA; London, UK, 2016.
66. Sijtsma, K. On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika* **2009**, *74*, 107–120. [CrossRef]
67. Schmitt, N. Uses and abuses of coefficient alpha. *Psychol. Assess.* **1996**, *8*, 350–353. [CrossRef]
68. Revelle, W.; Zinbarg, R.E. Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika* **2008**, *74*, 145–154. [CrossRef]
69. Zinbarg, R.E.; Revelle, W.; Yovel, I.; Li, W. Cronbach's α, Revelle's β, and Mcdonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* **2005**, *70*, 123–133. [CrossRef]
70. Zumbo, B.D.; Gadermann, A.M.; Zeisser, C. Ordinal versions of coefficients alpha and theta for Likert rating scales. *J. Mod. Appl. Stat. Methods* **2007**, *6*, 21–29. [CrossRef]
71. McDonald, R.P. *Factor Analysis and Related Methods*; Erlbaum: Hillsdale, MI, USA, 1985.
72. Gadermann, A.M.; Guhn, M.; Zumbo, B.D. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Pract. Assess. Res. Eval.* **2012**, *17*, 3.
73. Chalmers, R.P. On misconceptions and the limited usefulness of ordinal Alpha. *Educ. Psychol. Meas.* **2018**, *78*, 1056–1071. [CrossRef]

Paper IV

Zakariya, Y. F., Nilsen, H. K., Goodchild, S., & Bjørkestøl, K. (2020). Assessing first-year engineering students' pre-university mathematics knowledge: Preliminary validity results based on an item response theory model. *Journal of Technology and Science Education*, *10*(2), 259-270. doi:10.3926/jotse.1017

OmniaScience

Journal of Technology and Science Education
JOTSE, 2020 – 10(2): 259-270 – Online ISSN: 2013-6374 – Print ISSN: 2014-5349
https://doi.org/10.3926/jotse.1017

ASSESSING FIRST-YEAR ENGINEERING STUDENTS' PRE-UNIVERSITY MATHEMATICS KNOWLEDGE: PRELIMINARY VALIDITY RESULTS BASED ON AN ITEM RESPONSE THEORY MODEL

Yusuf F. Zakariya, H.K. Nilsen, Simon Goodchild, Kirsten Bjørkestøl

University of Agder, Department of Mathematical Sciences (Norway)

yusuf.zakariya@uia.no, hans.k.nilsen@uia.no, simon.goodchild@uia.no, kirsten.bjorkestol@uia.no

Received May 2020
Accepted June 2020

Abstract

The importance of students' prior knowledge to their current learning outcomes cannot be overemphasised. Students with adequate prior knowledge are better prepared for the current learning materials than those without the knowledge. However, assessment of engineering students' prior mathematics knowledge has been beset with a lack of uniformity in measuring instruments and inadequate validity studies. This study attempts to provide evidence of validity and reliability of a Norwegian national test of prior mathematics knowledge using an explanatory sequential mixed-methods approach. This approach involves use of an item response theory model followed by cognitive interviews of some students among 201 first-year engineering students that constitute the sample of the study. The findings confirm an acceptable construct validity for the test with reliable items and a high-reliability coefficient of .92 on the whole test. Mixed results are found on discrimination and difficulty indices of questions on the test with some questions having unacceptable discriminations and require improvement, some are easy, and some appear too tricky questions for students. Results from the cognitive interviews reveal the likely reasons for students' difficulty on some questions to be lack of proper understanding of the questions, text misreading, improper grasping of word-problem tasks, and unavailability of calculators. The findings underscore the significance of validity and reliability checks of test instruments and their effect on scoring and computing aggregate scores. The methodological approaches to validity and reliability checks in the present study can be applied to other national contexts.

Keywords – Prior knowledge, Item response theory, Mixed methods, Validity, Reliability.

To cite this article:

Zakariya, Y.F., Nilsen, H.K., Goodchild, S., & Bjørkestøl, K. (2020). Assessing first-year engineering students' pre-university mathematics knowledge: Preliminary validity results based on an item response theory model. *Journal of Technology and Science Education,* 10(2), 259-270. https://doi.org/10.3926/jotse.1017

----

1. Introduction

Students' knowledge before a teaching-learning activity has been reported in diverse fields of studies to exert enormous influence in facilitating proper understanding of current learning materials. Many psychological theories (e.g., self-efficacy theory) have acknowledged and emphasised this strong predictive role of prior knowledge on the current learning outcomes (Bandura, 1997; Marton & Booth, 1997). The correlation between prior academic achievement and students' performance in the presented tasks has been extensively reported in the literature. In a study of 60 first-year undergraduate students of accounting and economics reported by Duff (2004), prior academic achievement is found to correlate (r=.53) with academic performance positively, and it is the best among other predictors such as age and gender. This report is corroborated in a much larger sample longitudinal study in which prior academic achievement is also found to be the best, among other factors, in predicting 1,628 secondary school students' performance in reading and mathematics (Engerman & Bailey, 2006).

Furthermore, Ayán and García (2008) compare the efficacy of linear and logistic regression models in predicting 639 undergraduate students' performance, and both models favour prior academic achievement over other factors such gender and school location. In the same year, Hailikari, Nevgi and Komulainen (2008) conducted a special problem-solving mathematical assessment to determine students' prior knowledge in their study and its predictive power of academic performance. Their report shows that prior knowledge, coupled with previous academic success explained 55% of the variability observed in the performance of students on mathematics tasks. Similar results are also reported, elsewhere, (e.g., Casillas, Robbins, Allen, Kuo, Hanson & Schmeiser, 2012; Newman-Ford, Lloyd & Thomas, 2009; Richardson & Abraham, 2012).

Recently, a group of researchers Martin, Wilson, Liem and Ginns (2016) recorded mixed results on the prior knowledge predictive power of performance in a 2-year longitudinal study among university students. High school results as proxies for measuring prior knowledge correlate well with the performance at the beginning of their study while the ongoing semester course grades take the lead later. Though, this finding seems not contradictory to the earlier reported ones as both high school grades and the ongoing semester course grades still refer to prior academic achievement of the students in some sense. The findings reported by Aluko, Daniel, Oshodi, Aigbavboa and Abisuga (2018); Opstad, Bonesrønning and Fallan (2017) corroborate this point. Aluko et al. (2018) utilised more sophisticated statistical tools such as logistic regression and support vector machine learning to establish high correction between prior academic achievement and performance.

Despite the importance of prior knowledge and its correlation with students' performance, studies on psychometric properties of measures of engineering students' prior mathematics knowledge are scarce in the literature. As such, the primary purpose of the present study is to validate a prior knowledge of mathematics test (PKMT), owned by the Norwegian Mathematical Council, using an item response theory (IRT) model coupled with some cognitive interviews to extract detail information on likely reasons why some questions are challenging for students. The present study will not only provide empirical evidence for the validity of the PKMT but also offer pieces of advice to the Norwegian Mathematical Council towards an improvement of specific items on the test. Further, validation of the PKMT is also crucial for our ongoing relatively large-scale quantitative study on the contributions of prior mathematics knowledge, approaches to learning and self-efficacy on year-one engineering students' performance in mathematics at a Norwegian university. It is important to remark that the report presented in this article is preliminary and as such more studies are still ongoing in relating the scores of students on the prior knowledge of mathematics test to students' grades and other constructs.

The remaining part of the present article is arranged such that a conceptual framework is elucidated in the next section. The section was followed by another section where issues related to methodology, e.g., an overview of some specifics of the PKMT, sample of the study and procedure of data collection and analysis are presented. This is followed by a section where we present and discuss ensuing results from both the quantitative and the qualitative approaches to data analysis. The last section, before the reference list sheds more lights on the significant findings of the study, gives some concluding remarks and acknowledges the strengths and potential weaknesses of the study.

2. Conceptual Framework

2.1. The Conceptualisation of Prior Knowledge

There seems to be no agreement among educationists and psychologists on a definition of prior knowledge. Though, it is used to be captured as cognitive entry behaviour enshrined in Bloom's taxonomy (Bloom, 1976). The definition of Bloom's cognitive entry behaviour as "those prerequisite types of knowledge, skills, and competencies which are essential to the learning of a particular new task or set of tasks" (Bloom, 1976: page 122) has been criticised and considered outdated in some quarters (e.g. Dochy, De Rijdt & Dyck, 2002). In their review, Dochy et al. (2002) explicate many synonymous terms used to describe prior knowledge in the literature and consider their general interpretations to be "definitional snippets or vague statements" (Dochy et al., 2002: page 267). Thus, Dochy et al. (2002) propose and describe prior knowledge as:

> *The whole of a person's knowledge, which is as such dynamic in nature, is available before a certain learning task, is structured, can exist in multiple states (i.e. declarative, procedural and conditional knowledge), is both explicit and tacit in nature and contains conceptual and metacognitive knowledge components* (Dochy et al., 2002: page 267).

Another approach through which prior knowledge has been conceptualised is from an angle of domain-specific tasks or accomplishments. In this view, prior knowledge is seen as the level of knowledge related to a specific field being studied which varies distinctively depending on the relevance and the quality of the material currently under study (Dochy, 1996; Hailikari et al., 2008). Thus, prior knowledge in the present study refers to prior mathematics performance of students before they start their university education.

The notion of domain-specific prior knowledge seems to provide a basis for different indicators used in the literature to assess prior knowledge of the learners. There has been little coherence between various indicators used by educationists as proxies to quantify students' prior knowledge. This lack of uniformity can be linked to the type of studies, e.g. longitudinal (Engerman & Bailey, 2006), meta-analysis (Richardson & Abraham, 2012); students under study, e.g. university (Ayán & García, 2008), high school students (Casillas et al., 2012); the field of study, e.g. accounting (Duff 2004), mathematics (Hailikari et al., 2008), economics (Opstad et al., 2017), and architecture (Aluko et al., 2018). In several of these studies, researchers have used students' test scores on standardised tests, high school grades and entrance exams (Aluko et al., 2018; Casillas et al., 2012; Duff 2004; Newman-Ford et al., 2009) while others have used students previous semester/year grades (Ayán & García, 2008; Engerman & Bailey, 2006; Martin et al., 2016; Zakariya, 2016) or a special exam on problem-solving (Hailikari et al., 2008) to assess prior knowledge.

2.2. Study Setting

Students that are admitted into science and engineering courses at Norwegian universities have the freedom to choose between three routes and two endpoints for their mathematics studies at upper secondary schools (grades 11-13). The routes are practical mathematics (P-Mat) aiming at applications of mathematics, social science mathematics (S-Mat) and advanced mathematics for science and technology (R-Mat) and they can conclude their study of mathematics after two or three years at their upper secondary schools. The Norwegian Mathematical Council has consistently administered a prior knowledge of mathematics test (PKMT) to year-one university and college students since 1984. The PKMT aims to provide empirical evidence for monitoring of the basic knowledge of mathematics with a focus on undergraduate students following mathematics intensive programmes (e.g. engineering programmes) across universities and colleges in Norway. The PKMT is conducted every two years since 2001, and the latest was conducted in Autumn 2019. Prior to the year 2001, the test was conducted in 1984, 1986, 1999 and 2000. Accordingly, based on the results of the PKMT pieces of advice are offered by the Norwegian Mathematical Council to government agencies, Norwegian Research Council, universities, colleges, and other mathematics education stakeholders in Norway. However, it is apparent that some mathematics educators and researchers in Norway have reservations about the validity of the PKMT. It is the opinion of the authors that some of these reservations could be traced to a lack of validation studies on the instrument which has motivated the present study.

3. Methodology

3.1. Measure

The PKMT has two main parts. The first part contains background information about the students such as gender, age, some information about the highest mathematics content followed in upper secondary schools, and some items on attitudes towards mathematics. The second part is a 16-item test on basic mathematics tasks that are developed based on secondary school (grades 8-10) curriculum. Items 1 and 2 have three parts each, items 9 and 11 have two parts each, while other items have only one part each to make a total of 22 questions on the test. Questions 9a, 11a, 11b, 14 and 15 are standard multiple-choice questions while others are short open-ended questions. Before the commencement of the present study, the PKMT is administered using paper and pencil format. Thus, we independently digitalised the test and administered it online under classroom supervision. Coincidentally, the Norwegian Mathematical Council also shifted to digital PKMT in the 2019 administration of the test at the national level. The use of calculators is not allowed, and it takes 40 minutes to complete the test, including the time to complete background information. Sample questions of the PKMT are not included in the present article for confidentiality reasons. However, questions on the test can be categorised into five clusters: (a) basic operations of addition, multiplication, division and ordering of fractions and decimals; (b) simple percentages, ratio, proportion and average speed; (c) solving linear equations and inequalities including an application of Pythagoras theorem; (d) reading a Cartesian graph, slope of a straight line, similar triangles and volume of solid shapes; and (e) word problems on writing, interpreting and solving linear/simultaneous equations. Further, some of the questions are discussed in parts in a way that they are not identifiable during the presentation of some interview transcripts.

3.2. Participants

A total of 201 year-one engineering students in a Norwegian university including 34 females and 167 males took the PKMT in Autumn 2019. The average age of the students is 20.64 years, with a minimum of 17 years and a maximum of 36 years. Appropriate consents are sought from the Norwegian Centre for Research Data (NSD) as well as individual students who took parts in the test. The students are made to understand that taking part in the study is entirely voluntary and that their refusal to give consent will not in any way affect their grades. They are promised that utmost confidentiality will be ensured in dealing with their data and that no student is identifiable during and after the study. The data used for the present study are completely anonymous and are available upon request from the corresponding author.

3.3. Data Analysis

The collected data are initially scored dichotomously using 1 point for a correct answer and 0 point for a wrong. The scored data are analysed using a quantitative method. A two-parameter IRT model was used to investigate item parametrisations such as item discriminating and difficulty indices as well as item reliability of the test. An IRT model is a framework that characterises a relation between examinee's ability or latent trait as measured by a scale and the examinee's responses to each item on the scale (DeMars, 2010). IRT models can be one-parameter, two-parameter, three-parameter, unidimensional (i.e., items measure a common latent trait) and multidimensional (i.e., items measure separate clusters of a latent trait) depending on the complexity of the scale. The basic notion of the two-parameter IRT model is that a subject's probability of getting an item correct is a monotonic increasing function (e.g., an exponential function) of two sets of parameters: (a) the location (item difficulty) on the latent trait (in our case, prior mathematics knowledge) to be measured; and (b) the slope (item discrimination) of item response function (IRF) otherwise known as item characteristic curve (ICC). Equation 1 presents a mathematical representation of a two-parameter IRT model.

$$P(X_i = 1|\theta, a_i, b_i) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \quad (1)$$

Equation 1 shows the probability (P) that a student with latent variable θ (competence on the PKMT) answers an item ($X_i = 1$) correctly which has both item difficulty and discrimination indices of a_i and b_i respectively and *e* is an exponential function. The test scores on the PKMT is put on a metric determined by IRT model such that the group latent variable is normally distributed (mean = 0 and standard deviation = 1) with values ranging from -3.5 to +3.5. Each item discrimination index (a_i , i = 1, 2, …, 22) has the same metric as the latent variable (θ) with values ranging over the set of real numbers. It measures the extent to which an item discriminates between students of low and high ability on the PKMT. Items with negative or less than 0.20 a_i's have been recommended to be rejected while items with $0.20 \leq a_i$'s < 0.40 demonstrate appropriate discrimination and may be improved and items with a_i's ≥ 0.40 demonstrate good discrimination (DeMars, 2010; Ebel & Frisbie, 1991). However, depending on the sample size, item discrimination is not expected to be excessively high. Also, each item difficulty index (b_i , i = 1, 2, …, 22) is on the same metric as the latent variable (θ) with value range over the set of real numbers, and a practical range between -2 and 2 to avoid too easy or too tricky items on the test (DeMars, 2010). It gives information on the amount of the latent variable (θ) at which 50% of the students will get a correct score on each item. In as much as there seems to be no specific range of values to ascertain good difficulty index, empirical evidence has supported retaining items of the middle index of difficulty on the test (Ebel & Frisbie, 1991).

In line with unidimensionality assumption of IRT models (DeMars, 2010), a one-factor model of the PKMT with its 22 questions hypothesised to measure a common construct is evaluated using mean and variance adjusted unweighted least squares estimator with theta parametrisation (ULSMV-Theta). ULSMV-Theta is used because of its satisfactory performance and precision in estimating model parameters for a dichotomously scored IRT modelling in Mplus (Paek, Cui, Ozturk-Gubes & Yang, 2018). The model fit is assessed using multiple criteria. For an appropriate fit, we follow the recommendations of the ratio of chi-square value to the degree of freedom of less than 3coupled with a root mean square error of approximation (RMSEA) of less than .06 with non-significant 90% confidence interval (Brown, 2015), comparative fit and Tucker-Lewis indices (CFI and TLI) of greater than or close to .90 (Bentler, 1990). Further, we look at the significant level or otherwise of the factor loading of each of the items on the test. This is necessary to determine the contribution of these items to the test and to estimate each item reliability using standardised R-square values.

The qualitative method of data analysis takes the form of a cognitive interview. This interview was conducted to further probe and to determine the most likely reasons why some perceived too difficult questions based on the results of statistical analysis are not answered correctly. We rely on students' experience that voluntarily consented to take part in the interviews. In addition to the general consent to take part in the research project, special consent was requested from each student before the interview to audio record individual's utterances. The semi-structured cognitive interview was individually conducted in Norwegian using some leading questions with samples as follow:

> *If you, please take a look at this task, do you think this is a task you would have mastered?*
>
> *Do you have any idea on how to solve that one? Is it clear what they ask?*
>
> *What do you think is the reason why many students got this question incorrect?*
>
> *You get some calculations there, and you only have paper and pencil accessible, do you think that a calculator had been necessary for some students on this task?*

A total of seven students were interviewed, including six males and one female. Each interview lasted about 15 minutes, and the collected data were transcribed and translated into English. Selected results from these interviews are presented in the next section.

4. Results and Discussion

4.1. Results of Quantitative Analyses

Results from the analysis of a one-factor model of the PKMT with its 22 questions hypothesised to measure a common construct of students' prior mathematics knowledge are presented. Descriptive statistics of the analysed data as well as some initial parameters are shown in Table 1. The table shows the number of correct and incorrect responses of each item on the test, including the respective standardised factor loadings, R-square values, and the p-values.

Question	Number of correct responses	Number of incorrect responses	Factor loading	p-value	R-square	p-value
1A	173	28	.388	.001	.151	.091
1B	123	78	.472	< .001	.223	.006
1C	103	98	.444	< .001	.197	.006
2A	151	50	.454	< .001	.206	.017
2B	112	89	.562	< .001	.316	< .001
2C	85	116	.537	< .001	.288	< .001
3	64	137	.530	< .001	.281	< .001
4	116	85	.557	< .001	.310	< .001
5	143	58	.549	< .001	.301	< .001
6	128	73	.660	< .001	.436	< .001
7	113	88	.679	< .001	.461	< .001
8	133	68	.804	< .001	.646	< .001
9A	48	153	.169	.078	.029	.379
9B	19	182	.336	< .001	.113	.034
10	13	188	.553	< .001	.306	.001
11A	176	25	.891	< .001	.794	.003
11B	60	141	.350	< .001	.123	.033
12	82	119	.734	< .001	.539	< .001
13	58	143	.604	< .001	.365	< .001
14	86	115	.618	< .001	.382	< .001
15	125	76	.736	< .001	.542	< .001
16	67	134	.695	< .001	.483	< .001

Table 1. Descriptive statistics of the 22-item PKMT

The results presented in Table 1 reveal that Question 11A of the PKMT has the highest number of correct responses with 176 students got it correctly while Question 10 has the least number of correct responses with only 13 students got it correctly. All the item factor loadings are significant except for Question 9A, which has an insignificant factor loading of .169 ($p = .078$). These factor loadings reflect the strength at which each of the questions of the PKMT measures the purported prior mathematics knowledge the instrument is designed to measure. Thus, from this initial analysis, one can deduce that Question 9A has little or no substantial contribution to the instrument. Further, upon squaring each of these standardised factor loadings, a measure of variability (R-square) and reliability of each question on the PKMT was established. For instance, 31.6% and 79.5% variances of Question 2B and Question 11A, respectively, are explained by the latent construct of students' competence on the PKMT. And that these questions are reliable with significant reliability coefficients of .316 and .795, respectively. On the other hand, questions 1A, 2A, 9A, 9B and 11B have non-significant reliability coefficients of .151, .206, .029, .113, and .123, respectively, at $\alpha = .01$ level of significance. The reliability of the whole test was found to be .92 using a latent variable approach described in (Raykov, Dimitrov & Asparouhov, 2010; Raykov & Marcoulides, 2016). The goodness of fits statistics from the analysis of the one-factor PKMT model are presented in Table 2.

Model fit statistics	Values
Chi-square (χ^2)	
Value	272.892
df	209
χ^2/df	1.306
p-value	.002
CFI/TLI	
CFI	.903
TLI	.893
RMSEA	
Estimate	.039
90 per cent C.I.	.025 .051
Probability RMSEA <= .05	.927

Table 2. Selected goodness of fit indices of the one-factor PKMT model

The results presented in Table 2 show an appropriate fit of the evaluated one-factor model of the PKMT. The chi-square statistic seems a bit high and significant (p=.002). However, its ratio to the degree of freedom is less the recommended value of 3 for an acceptable model fit. Both the CFI and the TLI values are within the recommended values of an acceptable model fit (Bentler, 1990). The RMSEA is excellent with its value within the 90 per cent confidence interval, and its probability is not significant. This non-significant RMSEA probability shows that the model demonstrates a close fit of the data and that the hypothesis of not-close fit should be rejected (MacCallum, Browne & Sugawara, 1996). Thus, the overall fit statistics confirm that the hypothesised one-factor construct of prior mathematics knowledge exposed by the 22 questions is supported by empirical evidence. After establishing the model fit of the PKMT, we now turn to its item quality as explicated by item response theory parametrisation. The ensuing results on item discrimination and difficulty indices of each item on the PKMT as well as their respective p-values are presented in Table 3.

The results presented in Table 3 show that all the questions on the PKMT have acceptable item discrimination indices except for Question 9A (a_{9A} = 0.172, p = .087) and Question 11A (a_{11A} = 1.958, p = .223) which demonstrate too weak and too strong discriminations, respectively, among the students. The inference can be drawn from the non-significant estimates of the discrimination indices of these two questions. According to the classifications of item discrimination index by Ebel and Frisbie (1991), it can be inferred that our empirical evidence supports the removal of Question 9A and Question 11A from the test, Questions 9B and 11B have appropriate discriminating indices but can be improved upon, and all other questions have good discrimination indices. Further, it is also revealed in Table 3 that some questions demonstrate appropriate difficulty. At the same time, some questions demonstrate excessive item difficulty (i.e. too difficult questions), and other questions demonstrate weak difficulty (i.e. easy questions). For instance, questions 1C, 2B, 2C, 4, 7, 9A, 12, and 14 demonstrate appropriate difficulty with the non-significant estimates (p > .01) of their respective difficulty indices. Also, questions 1A, 1B, 2A, 5, 6, 8, 11A, and 15 are relatively easy questions depending on the absolute magnitude of their estimates while other questions, e.g. questions 3, 9B, 10, and 11B are difficult questions. Selected results of why students perceived some of these questions difficult are presented in the next section.

Question	Item discrimination	p-value	Item difficulty	p-value
1A	0.421	.004	-2.795	.002
1B	0.536	< .001	-0.602	.006
1C	0.495	< .001	-0.070	.725
2A	0.510	< .001	-1.494	< .001
2B	0.679	< .001	-0.256	.112
2C	0.637	< .001	0.362	.040
3	0.626	< .001	0.890	< .001
4	0.670	< .001	-0.350	.034
5	0.657	< .001	-1.016	< .001
6	0.877	< .001	-0.531	< .001
7	0.924	< .001	-0.231	.080
8	1.354	< .001	-0.518	< .001
9A	0.172	.087	4.198	.090
9B	0.357	< .001	3.908	< .001
10	0.664	< .001	2.741	< .001
11A	1.958	.223	-1.295	< .001
11B	0.374	< .001	1.509	.001
12	1.081	< .001	0.317	.012
13	0.758	< .001	0.923	< .001
14	0.786	< .001	0.294	.049
15	1.086	< .001	-0.422	.001
16	0.965	< .001	0.620	< .001

Table 3. IRT parameterisation of the PKMT

4.2. Results of Cognitive Interviews

To further probe why some questions are perceived difficult by the students, we interviewed some students to hear their views and suggestions for the improvement of such difficult questions. Results from the transcripts of interviews for Question 10 (this is a word-problem type question that requires the students to manipulate some percentages and give the final answer in decimal number) show that some students find it challenging to understand the question because of its practical and word-problem nature. Some of the reasons stated for getting the question incorrect by most students are lack of proper understanding of the question, text misreading, and unavailability of calculators. The students also think that provision of calculators during the test administration could improve their performance on such difficult tasks. For the reason that they are used to working on mathematical tasks with calculators lately, as mentioned by one of the students "*I would have thought about this for a while, I guess I had to because we are so used to use the calculator all the time*".

Similarly, when the interviewer asked the following questions about Question 9: What is difficult here? What makes this a bit difficult? If you could try to describe in words what makes it difficult to understand? Note: Question 9 is a word-problem type that requires the students to manipulate the purchase of oranges and bananas in kilogrammes using letters rather than numbers. One of the interviewees responded with the following answer:

> *No, it is more. I think it was difficult to understand. That the **a** stands for, I am more like I do not manage to deal with practical tasks after I began doing theoretical tasks […] This is a typical example of what I find difficult, that **a** is how many kilograms of oranges that you buy, and **b** is for bananas. What is 10**a** plus 15**b**? And then I become, like – actually I think I could have solved it if I had more time. If I had thought more about it. It is not how to solve it; it is more like I put a lot more energy into solving this task than this task [she points to task 3 which is on calculating the volume of a compound figure] because it is too much text and I become stressed, and I think back to the practical math that I had and that I did not like.*

It can be deduced from the excerpts of interview transcriptions for Question 9 that some students could not solve the problem correctly because of their inadequate reading comprehension, interpretations, and improper understanding of the word-problem task. Meanwhile, of the six difficult questions (3, 9B, 10, 11B, 13 and 16) identified in Table 3, only questions 3 and 11B are not posed in word problems. Thus, it can be inferred that the challenge with our students lies on their improper grasping of word-problem tasks which could stem from their preference for other types of mathematical tasks as evident in one of the student's response during the interview "*It is not how to solve it, it is more like I put a lot more energy into solving this task than this task* [she points to task 3 which is on calculating the volume of a compound figure] *because it is too much text and I become stressed, and I think back to the practical math that I had and that I did not like*". This finding conforms to the global trend of students' perceived difficulty of mathematical word-problem tasks at elementary, secondary and university levels (e.g., Vilenius‐Tuohimaa, Aunola & Nurmi, 2008; Zheng, Swanson & Marcoulides, 2011).

5. Conclusions

Prior mathematics knowledge of students has been identified as instrumental to the learning outcomes of current materials. Both theoretical and empirical evidence has been documented to support this claim (Bandura, 1997; Zakariya, 2016). However, proper assessment of students' prior mathematics knowledge has been beset with inconsistency in the available numerous measuring instruments and lack of validation studies. Attempts are made in the present study to validate a national test of prior mathematics knowledge of university students in Norway using mixed methods research design. The design involves the use of item response theory to provide psychometric properties of the test and cognitive interviews to probe plausible reasons why students find some questions challenging.

The findings of the present study provide empirical evidence for the construct validity of the Norwegian prior knowledge of mathematics test. In particular, our evaluation of a one-factor model shows that the test is measuring just a single latent variable (i.e. prior mathematics knowledge of students) that it is purported to measure. Further, it is also found that out of the 22 questions on the test only questions 1A, 2A, 9A, 9B and 11B demonstrate lack of acceptable reliability coefficients. However, the reliability coefficient of the whole test using latent variable approach is found to be very high (.92) which proves high internal consistency of the items on the test (Raykov et al., 2010). The latent variable approach is used to compute the reliability coefficient of PKMT because of its reported excellent performance over the popular Cronbach's alpha and Kuder-Richardson formula 20 (e.g., Raykov et al., 2010; Raykov & Marcoulides, 2016). In as much as most of the reviewed literature in the present study (e.g., Hailikari et al., 2008; Newman‐Ford et al., 2009) do not report reliability coefficients of their measures of prior academic knowledge, the reliability coefficient of the PKMT is higher than the one reported by Lee and Chen (2009) but slightly lower than the Kuder-Richardson coefficient reported by Casillas et al. (2012).

The findings of the present study also show that questions on the PKMT are at different levels of difficulty and variant discriminations between students of low and high competence in the prior mathematics knowledge test. These findings have several implications on the validity and reliability of aggregate scores of the test and other analyses (e.g. means comparisons between universities and previous years) usually presented by the Norwegian Mathematical Council. For instance, the assignment of a score of 1 point to an easy and poorly discriminating item, e.g., Question 11A and to a challenging and good discriminating item, e.g., Question 10 may bias the aggregate scores of students with low ability upward on the test and reduce the aggregate scores of highly competent students. This kind of bias in aggregate scores is a threat to the validity of the test and a typical disadvantage of using classical test theory approach in scoring tests (DeMars, 2010). Thus, we urge the Norwegian Mathematical Council to use item response theory which can incorporate the test item difficulty and discrimination indices in the scoring process such that more valid aggregate scores can be obtained. Moreover, of course, a more reliable mean score comparison can be made. Further, compelling evidence is also provided in the findings of the present study that suggests the removal of or at least improvement in item wordings and presentation of questions 9A, 9B, 11A and 11B on the test.

Moreover, six out of the 22 questions of PKMT are also found to be very difficult for students to answer correctly. Empirical evidence from cognitive interviews of some students who took part in the test reveals potential reasons why these questions are perceived difficult. Some of the ascribed causes of poor performance on these questions are lack of proper understanding of the question, text misreading, improper grasping of word-problem mathematical tasks, and unavailability of calculators. Given that low performance on word-problem tasks is not peculiar to Norwegian engineering students (e.g., Vilenius-Tuohimaa et al., 2008), we recommend innovative teaching and learning strategies to alleviate these problems. Such strategies can be the use of modelling activities, problem-based learning, and so on (Greer, 1997; Zakariya, Ibrahim & Adisa, 2016) to foster understanding and interpretation of word-problem mathematical tasks. The Norwegian Mathematical Council may also consider the introduction of calculators in subsequent PKMT administrations. Finally, this section is concluded by acknowledging some strengths and potential weaknesses of the findings of this study.

6. Strengths and Potential Weaknesses of this Study

A strength of this study lies in the use of explanatory sequential mixed-methods approach to data analysis (Bryman, 2016) that involves a robust quantitative analysis procedure in terms of an IRT followed by some cognitive interviews. The interviews avail us the opportunity to look at the data beyond statistical analyses and provide a more elaborate description of the phenomenon. Another strength of this study encompasses a relatively large data set of 201 engineering students used in the present study. The large sample involved is a potential for generalisation of our findings, especially now that such large-scale study is scarce in mathematics education research. However, a potential limitation of the present study could stem from a lack of external validity of the PKMT. There was no independently measured variable such as students' grades, and grade point average through which the predictive validity of the PKMT can be confirmed. We did not investigate the content validity of the test items as we lack the permission to do so. Instead, our findings only provide evidence for its psychometric property. Also, the restriction of the sample of the study to a Norwegian university and only engineering students might, in a way, limit the generalisation of our findings. Thus, we recommend the replications of the present study in more substantial and more diverse university student populations. Despite these limitations, our study does provide potential cues on the construct validity, reliability, and item quality of the PKMT which will be useful to Norwegian Mathematical Council, researchers, and other stakeholders in mathematics education. The methodology adopted in the present study can also be applied in other national contexts to investigate the validity of their measures.

Acknowledgements

The authors appreciate the kind gesture of the Norwegian National Mathematical Council for permitting us to use their test.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest for the research, authorship, and publication of this article.

Funding

The authors received no financial support for the research and authorship of this article. Meanwhile, we appreciate the University of Agder library for funding the article processing charge of this article.

References

Aluko, R.O., Daniel, E.I., Oshodi, O.S., Aigbavboa, C.O., & Abisuga, A.O. (2018). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering, Design and Technology*, 16(3), 385-397. https://doi.org/10.1108/jedt-08-2017-0081

Ayán, M.N.R., & García, M.T.C. (2008). Prediction of university students' academic achievement by linear and logistic models. *The Spanish Journal of Psychology*, 11(1), 275-288. https://doi.org/10.1017/s1138741600004315

Bandura, A. (1997). S*elf-efficacy: The exercise of control*. New York: W.H. Freeman and Company.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. https://doi.org/10.1037/0033-2909.107.2.238

Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. New York, London: The Guilford Press.

Bryman, A. (2016). *Social research methods*. Oxford, United Kingdom: Oxford University Press.

Casillas, A., Robbins, S., Allen, J., Kuo, Y.L., Hanson, M.A., & Schmeiser, C. (2012). Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology*, 104(2), 407-420. https://doi.org/10.1037/a0027180

DeMars, C. (2010). *Item response theory: Understanding statistics measurement*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195377033.001.0001

Dochy, F. (1996). Assessment of domain-specific and domain-transcending prior knowledge: Entry assessment and the use of profile analysis. In Birenbaum M., & Dochy, F. (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge. Evaluation in Education and Human Services* (42, 227-264).Dordrecht: Springer. https://doi.org/10.1007/978-94-011-0657-3_9

Dochy, F., De Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning. How far have we progressed since Bloom? Implications for educational practice and teaching. *Active learning in higher education*, 3(3), 265-284. https://doi.org/10.1177/1469787402003003006

Duff, A. (2004). Understanding academic performance and progression of first-year accounting and business economics undergraduates: The role of approaches to learning and prior academic achievement. *Accounting Education*, 13(4), 409-430. https://doi.org/10.1080/0963928042000306800

Ebel, R., & Frisbie, D.A. (1991). *Essentials of educational measurement*. New Delhi: Prentice - Hall of India.

Engerman, K., & Bailey, U.J.O. (2006). Family decision-making style, peer group affiliation and prior academic achievement as predictor of the academic achievement of African American students. *The Journal of Negro Education*, 75(3), 443-457.

Greer, B. (1997). Modelling reality in mathematics classrooms: the case of word problems. *Learning and Instruction*, 7(4), 293-307. https://doi.org/10.1016/S0959-4752(97)00006-6

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: A structural model. *Educational Psychology*, 28(1), 59-71. https://doi.org/10.1080/01443410701413753

Lee, C.Y., & Chen, M.P. (2009). A computer game as a context for non-routine mathematical problem solving: the effects of type of question prompt and level of prior knowledge. *Computers & Education*, 52(3), 530-542. https://doi.org/10.1016/j.compedu.2008.10.008

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods,* 1(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Martin, A.J., Wilson, R., Liem, G.A.D., & Ginns, P. (2016). Academic momentum at university/college: Exploring the roles of prior learning, life experience, and ongoing performance in academic achievement across time. *The Journal of Higher Education*, 84(5), 640-674. https://doi.org/10.1353/jhe.2013.0029

Marton, F., & Booth, S.A. (1997). *Learning and awareness*. Hillsdale, NJ: Lawrence Erlbaum.

Newman-Ford, L., Lloyd, S., & Thomas, S. (2009). An investigation in the effects of gender, prior academic achievement, place of residence, age and attendance on first-year undergraduate attainment. *Journal of Applied Research in Higher Education*, 1(1), 14-28. https://doi.org/10.1108/17581184200800002

Opstad, L., Bonesrønning, H., & Fallan, L. (2017). Tar vi opp de rette studentene ved økonomisk-administrative studier? *Samfunnsøkonomen,* 1, 21-29.

Paek, I., Cui, M., Ozturk-Gubes, N., & Yang, Y. (2018). Estimation of an IRT model by Mplus for dichotomously scored responses under different estimation methods. *Educational and Psychological Measurement*, 78(4), 569-588. https://doi.org/10.1177/0013164417715738

Raykov, T., Dimitrov, D.M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 265-279. https://doi.org/10.1080/10705511003659417

Raykov, T., & Marcoulides, G.A. (2016). Scale reliability evaluation under multiple assumption violations. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(2), 302-313. https://doi.org/10.1080/10705511.2014.938597

Richardson, M., & Abraham, C. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387. https://doi.org/10.1037/a0026838

Vilenius-Tuohimaa, P.M., Aunola, K., & Nurmi, J.E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, 28(4), 409-426. https://doi.org/10.1080/01443410701708228

Zakariya, Y.F. (2016). Investigating validity of Math 105 as prerequisite to Math 201 among undergraduate students, Nigeria. *International Journal of Instruction*, 9(1), 107-118. https://doi.org/10.12973/iji.2016.919a

Zakariya, Y.F., Ibrahim, M.O., & Adisa, L.O. (2016). Impacts of problem-based learning on performance and retention in mathematics among junior secondary school students in Sabon-Gari area of Kaduna state. *International Journal for Innovative Research in Multidisciplinary Field*, 2(9), 42-47.

Zheng, X., Swanson, H.L., & Marcoulides, G.A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology*, 110(4), 481-498. https://doi.org/10.1016/j.jecp.2011.06.001

Published by OmniaScience (www.omniascience.com)

Journal of Technology and Science Education, 2020 (www.jotse.org)

Paper V

Zakariya, Y. F., Nilsen, H. K., Bjørkestøl, K., & Goodchild, S. (2020). Impact of attitude on approaches to learning mathematics: a structural equation modeling approach. In T. Hausberger, M. Bosch & F. Chellougthi (Eds.), *Proceedings of the Third Conference of the International Network for Didactic Research in University Mathematics (INDRUM 2020, 12-19 September 2020)* (pp. 268 - 277). Bizerte, Tunisia: University of Carthage and INDRUM.

Impact of attitude on approaches to learning mathematics: a structural equation modelling approach

Yusuf F. Zakariya[1], H. K. Nilsen[1], Kirsten Bjørkestøl[1], and Simon Goodchild[1]

[1]University of Agder in Kristiansand, Faculty of Engineering and Science, Norway, yusuf.zakariya@uia.no

This study aims at validating an attitude subscale of a national mathematics test that has been repeatedly used for over two decades and to expose the relations between students' attitude towards mathematics and their approaches to learning mathematics. A sample of 196 year-one engineering students completed two survey instruments used for the study. Using a structural equation modelling approach, empirical evidence of construct validity, discriminant validity and reliability were found for the attitude subscale. Further, it was also found that students' attitude towards mathematics had a substantial positive impact on deep approaches to learning and a substantial negative impact on surface approaches to learning. These findings could be of help to university teachers and other stakeholders in designing appropriate interventions to support the students.

Keywords: students' practices, deep approach, surface approach, attitude, structural equation modelling.

INTRODUCTION

Approaches to learning in higher education are part of students' practices that have a considerable effect on learning outcomes. Students who care for every detail in their course content with the intent to achieve conceptual understanding (deep approach) are more likely to perform better than others who only rely on memorization of key points (surface approach). Approaches to learning have been conceptualized to include "predispositions adopted by an individual when presented with learning materials and strategies used to process the learning contents" (Zakariya, Bjørkestøl, Nilsen, Goodchild, & Lorås, 2020). It is an essential factor in students' practices that has received increased attention in recent times. Perhaps, as a result of international campaigns on aligning university education towards developing learners' deep approaches that will enable them to navigate easily through an increasingly changing society.

Several empirical studies have been reported on the factors that encourage or discourage the adoption of either deep or surface approaches to learning. One of these studies is a critical review by Baeten, Kyndt, Struyven, and Dochy (2010). Therein, a total of 118 empirical studies were reviewed, and the results can be summarized as follows: satisfaction with course quality, big five personality traits except for neuroticism, and emotional stability are some of the factors that stimulate adoption of the deep approaches to learning. It was also found that students that experience intrinsic motivation, and who are self-efficacious and self-confident are most likely to adopt deep approaches to learning. In a follow-up quasi-experimental study Baeten, Struyven, and Dochy (2013) investigated the contribution of some teaching methods on students' approaches to learning. They found that adoption of deep approaches to learning decreases among the participants in a lecture-based group while they remain stable in a student-centred learning environment over a period.

More so, Von Stumm and Furnham (2012) conducted an empirical study involving 579 psychology and computer science undergraduate students on relations between approaches to learning, personality, intelligence and intellectual engagement. It was found that deep approaches to learning strongly related to intellectual engagement while personality and intelligence explained 25% variability in surface approaches to learning among the subjects of their study. In an attempt to unravel the interwoven bond between critical thinking, self-efficacy and learning approaches, Hyytinen, Toom, and Postareff (2018) conducted an empirical study involving 92 science education undergraduate students in Finland. Their results showed that students with high self-efficacy also adopt deep approaches to learning. Some researchers have studied relations between approaches to learning and other factors in domain-specific contexts. For instance, Mji (2000) found that there was a strong relationship between students' different conceptions of mathematics and their approaches to learning the subject.

Despite the importance of approaches to learning and its relations with some affective constructs, e.g. self-efficacy there are few studies on its relations with attitudes of students. One of the relatively recent studies on this topic is the report by Alkhateeb and Hammoudi (2006) on the relations between attitude towards mathematics and students' approaches to learning. In their study, students with a positive attitude towards mathematics were identified with deep approaches to learning while those with a negative attitude towards mathematics were identified with surface approaches to learning approaches. However, their study had some methodological issues such as the use of regression analysis to examine the relations between these constructs, given that the regression analysis does not account for measurement errors in the predictor variable(s). Another methodological issue in their study involved the use of mean scores derived from item parcelling of ordinal variables which could lead to biased results because of violation of multiple assumptions, e.g. tau-equivalent, and normal distribution (Zakariya, 2020).

Thus, the present study was motivated by the sparsity of studies on the relationship between attitude towards mathematics and approaches to learning coupled with some methodological issues observed in available studies (e.g., Alkhateeb & Hammoudi, 2006). Further, to the best of our knowledge, there was no validation study on the attitude subscale of the Norwegian national mathematics test for the past fifteen years. The national mathematics test is a test that is conducted every two years and designed to assess pre-university knowledge of mathematics of year-one undergraduate students across universities in Norway. The validity of this test is essential to ensure the test measures what is purported to measure, which will facilitate more accurate interpretations of its ensuing results. Therefore, the primary purposes of the present study are to use a structural equation modelling approach to (a) validate the attitude subscale of the Norwegian national mathematics test; (b) expose the impact of attitude towards mathematics on students' learning approaches. The use of the structural equation modelling approach will avail us an opportunity of taking of care of the two methodological issues involved in the use regression analysis that is typically used in the literature (e.g., Alkhateeb & Hammoudi, 2006). In the next section, a conceptual framework coupled with a theoretical perspective that justifies the rationale for finding the relations between these constructs is discussed.

CONCEPTUAL FRAMEWORK

A theoretical structure that could be used to justify the relations between attitude towards mathematics and approaches to learning is social cognitive theory. This theory sees an individual's behavioural changes as consistently being regulated and modified by interacting with social factors in the environment whose feedback influences the next actions and outcomes (Bandura, 2001). Central to this theory is the concept of reciprocal determinism that postulates a dynamic relationship between personal, behavioural, and environmental determinants (Bandura, 2012). Even though both the attitude towards mathematics and approaches to learning are personal factors, it is presumed that the dynamic relationship between the determinants (personal, behavioural, and environmental) can be extrapolated to within the personal determinants (cognitive, affective and biological factors). As such, a causal relationship between attitude towards mathematics and approaches to learning can be theoretically postulated. Empirical evidence has shown that students' attitude towards learning mathematics is greatly influenced by consistent interactions with teachers, peer groups and parents (e.g., Davadas & Lay, 2017). In other words, students whose teachers are efficacious, motivate them to learn, give positive feedback, maintain good teacher-student relations are more likely to develop a positive attitude towards mathematics. This, in turn, influences their approaches to learning the subject.

Several attempts have been made to conceptualize and operationalize both attitude towards mathematics and approaches to learning. Attitude towards mathematics has been conceptualized to include appraisal, valuation and enjoyment of mathematics (Zakariya, 2017). It is a construct whose multifaceted nature has influenced, to a great extent, the development of its measuring instruments (e.g., Palacios, Arias, & Arias, 2013; Zakariya, 2017). Some of these instruments have contributed significantly to the measurement of this construct as well as in relating it to other constructs from quantitative research perspectives. However, for the purpose of this study, a 5-item unidimensional attitude scale which is part of the national mathematics test in Norway, was selected. Our choice of this scale was prompted by two factors: (a) availability in the Norwegian language; (b) our quest to provide construct and discriminant validity which is lacking in the literature.

In addition, the "revised two-factor study process questionnaire" (R-SPQ-2F) has been identified as one of the best instruments for measuring students' approaches to learning

(López-Aguado & Gutiérrez-Provecho, 2018). R-SPQ-2F was chosen for the present study because of its high psychometric properties, a small number of items and ease of score interpretations. Further, Norwegian validations of this instrument have been undertaken (e.g., Zakariya, 2019; Zakariya et al., 2020). The Norwegian version has ten items on deep subscale and nine items on surface subscale with evidence of construct validity, discriminant validity, and internal consistency of its items.

Based on the postulates of the social cognitive theory coupled with previous literature, the two hypotheses of the present study are stated as follows, while Figure 1 depicts these hypothesized relations:

> (H01) There are substantial positive impacts of attitude towards mathematics on deep approaches to learning.
>
> (H02) There are substantial negative impacts of attitude towards mathematics on surface approaches to learning.

Figure 1 shows hypothesized impact of attitude measured by five items (att01 – att05) on both deep and surface approaches each measured by ten items and nine items respectively with an error correlation (indicated by the double-headed arrow) between deep and surface approaches. The plus (+) and minus (-) signs indicate the hypothesized positive and negative impacts of attitude on deep and surface approaches, respectively.

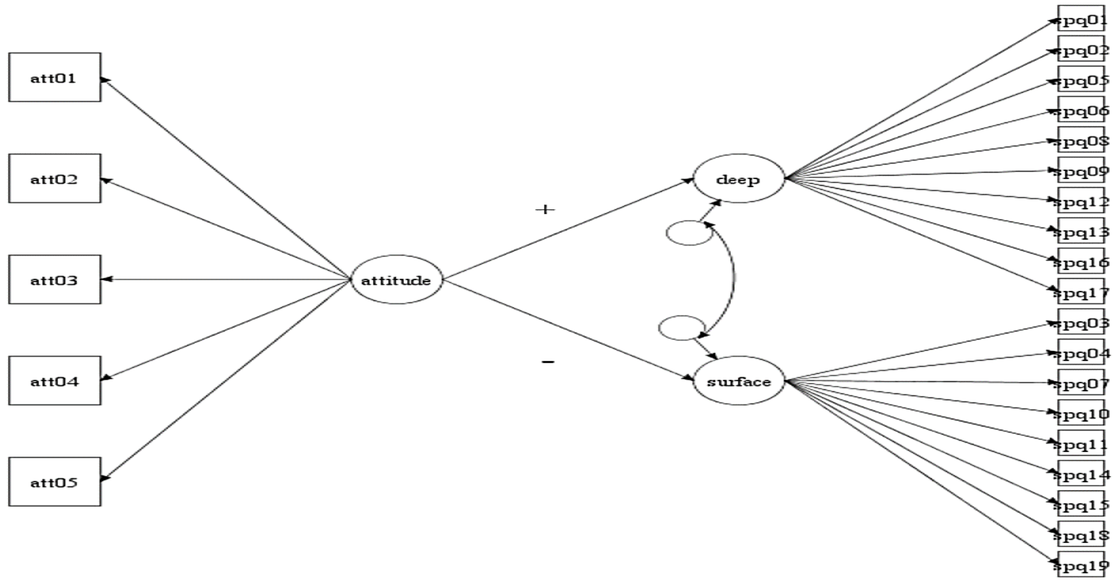


Figure 1: A hypothesized model of the relations between attitude towards mathematics and approaches to learning mathematics

METHOD

SAMPLE AND MEASURES

The sample for this study was made up of 196 year-one engineering students, including 34 females and 162 males with an average age of 24.64 years. Two online survey instruments were completed by the students, including R-SPQ-2F (Norwegian version) and attitude towards mathematics scale (AtMS). R-SPQ-2F is a 19-item questionnaire in which respondents rated their level of agreement from (1) '*never* or *only rarely* true of me' to (5) '*always* or *almost always* true of me' to statements like "I test myself on important topics until I understand them completely" (deep approach), "I see no point in learning material which is not likely to be in the examination" (surface approach), etc. On the other hand, AtMS is a 5-item scale in which respondents rated their level of agreements from (1) '*strongly disagree*' to (4) '*strongly agree*' to statements like "I work with mathematics because I like it", and "I'm interested in what I learn in math".

DATA ANALYSIS

The analyses proceeded in two stages. Stage one involved fitting a measurement model to examine the construct validity and unidimensionality of the AtMS. In this stage, AtMS data were screened for outliers, normality assumption, skewness, and kurtosis. It was found that AtMS contained excess kurtosis (absolute value > 2) and both Kolmogorov-Smirnov's and Shapiro-Wilk's tests were significant for each item which showed that the data were not normally distributed. Thus, weighted least square mean and variance adjusted (WLSMV) estimator was used for the confirmatory factor analysis as it is robust enough to perform well under violation of multiple assumptions (Suh, 2015; Zakariya, Goodchild, Bjørkestøl, & Nilsen, 2019). Further, both the item and scale reliability indices of AtMS were investigated using latent factor approach as opposed to the Cronbach alpha coefficient.

Analyses in stage two involved validating a structural model that explains the relations between attitude towards mathematics and approaches to learning. It consisted of evaluating the model and conducting exploratory post hoc analysis for its improvement. The structural equation modelling approach was used to either confirm or falsify the causal hypothesized relations between attitude towards mathematics and approaches to learning without claiming outright causation between the constructs. Model fits were assessed based on a combination of criteria as proposed in literature which includes: χ^2 ratio to the degree of freedom (df) less than 3, significant estimated parameters, comparative fit index (CFI), Tucker-Lewis index (TLI) close to or ≥ .95, root mean square error of approximation (RMSEA) ≤ .06, and standardized root mean square residual (SRMR) ≤ .08 (Chen, 2007; Hu & Bentler, 1999). All the analyses were performed using Mplus 8.3 software, and the results are presented in the next section.

RESULTS

STAGE ONE: MEASUREMENT MODEL AND RELIABILITY

A one-factor model was evaluated for the measure of attitude towards mathematics, and the results are presented in Table 1.

GOF indices	Model 1	Model 2
χ^2-value	84.078	3.562

df	5	3
p-value	< .001	.313
χ^2/df	16.816	1.187
CFI	.927	.999
TLI	.854	.998
RMSEA (90% CI)	.284 (.233 - .339)	.031 (< .001 - .128)
CFit	< .001	.510
SRMR value	.076	.038

Table 1: Selected goodness of fit indices (GOF) for measurement 1-factor model of AtMS
Note. CFit: close fit (i.e. probability of RMSEA <= .05)

The results of the selected goodness of fit (GOF) indices, presented in Table 1 (Model 1) showed an appropriate fit of the one-factor model AtMS. However, the significant chi-square value, its ratio to df > 3, and the low value of TLI, suggest that the model can be improved. Thus, a post hoc analysis was conducted using suggestions from modification indices. On this basis, two error covariances were included in the model between item 02 and item 04 as well as between item 01 and item 05. These resulted in a significant improvement in the model (Model 2) as indicated by the significant chi-square difference test statistics with Satorra-Bentler correction $\Delta\chi^2_{[2]} = 80.516, p < .001$. The model chi-square value is no longer significant, which is expected and its ratio to df < 3, CFI, TLI, RMSEA and SRMR, all now within the recommended ranges. These suggest an excellent fit of Model 2. All the factor loadings are found to be significant, and all the items are reliable with an ordinal coefficient alpha of .78 on the whole measuring instrument.

STAGE TWO: STRUCTURAL MODEL

In an attempt to test hypotheses one and two, we evaluated two structural models. The first model (Model 3) concerns the impact of attitude towards mathematics on the two dimensions of approaches to learning the subject. The second model (Model 4) concerns the final improvement of Model 3 through post hoc analyses. Selected GOF indices of these models are presented in Table 2, while Figure 2 displays standardized estimates of factor loadings, regression weights, variance explained, etc.

GOF indices	Model 3	Model 4
χ^2-value	425.784	301.876
df	247	204
p-value	< .001	< .001
χ^2/df	1.724	1.480

CFI	.913	.953
TLI	.902	.947
RMSEA (90% CI)	.061 (.051 - .070)	.049 (.037 - .061)
CFit	.037	.518
SRMR value	.070	.059

Table 2: Selected goodness of fit indices for measurement 1-factor model of AtMS

The results in Table 2 (Model 3) showed an appropriate fit of the model except that both CFI and TLI are relatively low and close fit probability assessment of RMSEA was significant which implies the model is not close enough to the data. As the first step in post hoc analysis of the structural equation modelling approach, we scanned through the estimates and discovered that item 4 and item 10 of the surface approach subscale of R-SPQ-2F had non-significant factor loadings. These items were deleted from the model, and the resulting model improved significantly as indicated by the significant chi-square difference test statistics with Satorra-Bentler correction $\Delta\chi^2_{[43]} = 123.908, p < .001$. The results, as presented in Table 2 (Model 4) suggest an excellent fit of the model. Figure 2 gives more detail on the parameter estimates of Model 4.

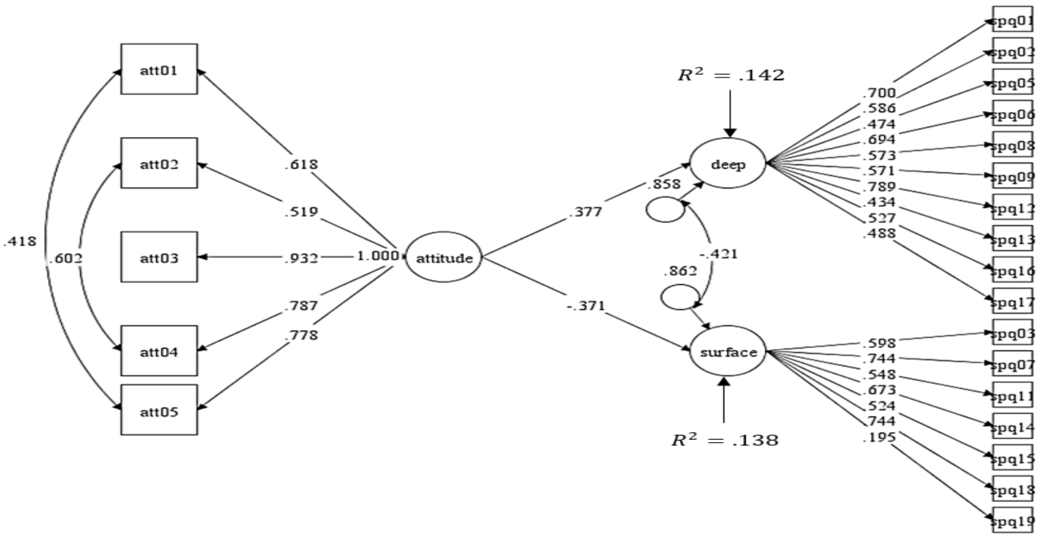


Figure 2: Validated structural model of the relations between attitude towards mathematics and approaches to learning mathematics

The illustrated results by Figure 2 show that there is a significant positive impact of attitude towards mathematics on deep approaches to learning (β= .377, p < .05) and a significant negative impact on the surface approaches to learning (β= -.371, p < .05) which confirm hypothesis one (H01) and hypothesis two (H02) respectively. These findings could be interpreted to mean that students who have a high (positive) attitude towards learning mathematics are more likely to adopt deep approaches to learning the subject. On the other hand, students who have a low (negative) attitude towards mathematics are more likely to adopt surface approaches to learning the subject. It is also revealed in Figure 2 that attitude towards mathematics explained 14.2% and 13.8% variances in predicting deep and surface approaches, respectively. These percentages of explained variances appear low. However, they are statistically significant. The low percentages of explained variances in deep and surface approaches are suggestive of the presence of other factors that are not captured in the present study and yet influence the adoption of students' learning approaches. In the next section, we present a brief discussion of the significant findings.

DISCUSSION

Attempts are made in the present study to provide empirical evidence for construct and discriminant validity of a 5-item attitude subscale of the Norwegian national mathematics test and to expose the impact of attitude towards mathematics on students' learning approaches. The attitude subscale was found to be unidimensional, and possesses construct validity, it discriminates cleanly between two approaches to learning and it has high internal item consistency with an ordinal coefficient of .78. However, this validity evidence was achieved after accounting for two error covariances between item 02: "I work with math because I like it" and item 04: "I'm interested in what I learn in math" as well as between item 01: "making an effort in math is important because it will help me in work I will be doing later" and item 05: "mathematics is an important subject for me because I need it when I want to study further".

It is important to remark that the error covariances appear to make sense conceptually since both item 02 and item 04 seem to capture intrinsic motivation part of attitude and item 01 and item 05 seem to capture usefulness of mathematics part of attitude. This finding corroborates other studies that have reported multidimensional attitude scales (Palacios et al., 2013). Further, the reliability coefficient of the AtMS ($\alpha = .78$) is higher than that of the perception of utility subscale ($\alpha = .679$) reported in (Palacios et al., 2013) and that of the usefulness of mathematics subscale ($\alpha = .75$) reported in (Zakariya, 2017) even though the final reliability coefficients of the whole scales reported in the two previous studies are higher than $\alpha = .78$ that was found for the AtMS.

Another important finding of the present study is the substantial positive impact of attitude towards mathematics on the deep approaches to learning as well as the substantial negative impact on the surface approaches to learning. These findings, on the one hand, suggest that year-one engineering students who enjoy mathematics, who are interested in the subject and recognize the utility of mathematics to their future studies are more likely to adopt deep approaches to learning the subject. On the other hand, the findings suggest that year-one engineering students who find mathematics less enjoyable and struggle to discover its relevance to their future studies may tend to adopt surface approaches to learning the subject. These findings agree with the report by Alkhateeb and Hammoudi (2006) and partly support some reported results by García, Rodríguez, Betts, Areces, and González-Castro (2016). More importantly, we do not claim outright causal relations between these constructs. However, our results have only provided tentative empirical evidence that confirms our hypothesized causal relations between engineering students' attitude towards mathematics and approaches to learning. Future replication studies are recommended to confirm these findings in independent samples. Finally, it is hoped that the findings of this study have shed some light on a general understanding of the causal relations between the attitude of students towards mathematics and their approaches to learning the subject. This could be of help to university teachers and other stakeholders in designing appropriate interventions to support the students.

REFERENCES

Alkhateeb, H. M., & Hammoudi, L. (2006). Attitudes toward and approaches to learning first-year university mathematics. *Perceptual and Motor Skills, 103*, 115-120. doi:10.2466/pms.103.1.115-120

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243-260. doi:10.1016/j.edurev.2010.06.001

Baeten, M., Struyven, K., & Dochy, F. (2013). Student-centred teaching methods: Can they optimise students' approaches to learning in professional higher education? *Studies in Educational Evaluation, 39*(1), 14-22. doi:10.1016/j.stueduc.2012.11.001

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1-26. doi:10.1146/annurev.psych.52.1.1

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, *38*(1), 9-44. doi:10.1177/0149206311410606

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. doi:10.1080/10705510701301834

Davadas, S. D., & Lay, Y. F. (2017). Factors affecting students' attitude toward mathematics: A structural equation modeling approach. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(1), 517-529. doi:10.12973/ejmste/80356

García, T., Rodríguez, C., Betts, L., Areces, D., & González-Castro, P. (2016). How affective-motivational variables and approaches to learning predict mathematics achievement in upper elementary levels. *Learning and Individual Differences, 49*, 25-31. doi:10.1016/j.lindif.2016.05.021

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Hyytinen, H., Toom, A., & Postareff, L. (2018). Unraveling the complex relationship in critical thinking, approaches to learning and self-efficacy beliefs among first-

year educational science students. *Learning and Individual Differences*, *67*, 132-142. doi:10.1016/j.lindif.2018.08.004
López-Aguado, M., & Gutiérrez-Provecho, L. (2018). Checking the underlying structure of R-SPQ-2F using covariance structure analysis / Comprobación de la estructura subyacente del R-SPQ-2F mediante análisis de estructura de covarianza. *Cultura y Educación/ Culture and Education, 30*(1), 105-141. doi:10.1080/11356405.2017.1416787
Mji, A. (2000). Conceptions of and approaches to learning mathematics. *Psychological Reports, 87*, 971-972. doi:10.2466/pr0.2000.87.3.971
Palacios, A., Arias, V., & Arias, B. (2013). Attitudes Towards Mathematics: Construction and Validation of a Measurement Instrument // Las actitudes hacia las matemáticas: Construcción y validación de un instrumento para su medida. *Revista de Psicodidactica / Journal of Psychodidactics, 19*(1), 67-91. doi:10.1387/RevPsicodidact.8961
Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(4), 568-580. doi:10.1080/10705511.2014.937669
Von Stumm, S., & Furnham, A. F. (2012). Learning approaches: Associations with typical intellectual engagement, intelligence and the big five. *Personality and Individual Differences, 53*(5), 720-723. doi:10.1016/j.paid.2012.05.014
Zakariya, Y. F. (2017). Development of attitudes towards mathematics scale (ATMS) using Nigerian data – Factor analysis as a determinant of attitude subcategories. *International Journal of Progressive Education, 13*(2), 74-84.
Zakariya, Y. F. (2019). Study approaches in higher education mathematics: Investigating the statistical behaviour of an instrument translated into Norwegian. *Education Sciences, 9*(3), 191. doi:10.3390/educsci9030191
Zakariya, Y. F. (2020). Investigating some construct validity threats to TALIS 2018 teacher job satisfaction scale: Implications for social science researchers and practitioners. *Social Sciences, 9*(4). doi:10.3390/socsci9040038
Zakariya, Y. F., Bjørkestøl, K., Nilsen, H. K., Goodchild, S., & Lorås, M. (2020). University students' learning approaches: an adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Education Evaluation*, 100816. doi:10.1016/j.stueduc.2019.100816
Zakariya, Y. F., Goodchild, S., Bjørkestøl, K., & Nilsen, H. K. (2019). Calculus self-efficacy inventory: Its development and relationship with approaches to learning. *Education Sciences, 9*(3), 170. doi:10.3390/educsci9030170

Paper VI

Zakariya, Y. F., Nilsen, H. K., Goodchild, S., & Bjørkestøl, K. (2020). Self-efficacy and approaches to learning mathematics among engineering students: Empirical evidence for potential causal relations. *International Journal of Mathematical Education in Science and Technology*, 1-15. doi: 10.1080/0020739X.2020.1783006

Taylor & Francis
Taylor & Francis Group

International Journal of Mathematical Education in Science and Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmes20

Self-efficacy and approaches to learning mathematics among engineering students: empirical evidence for potential causal relations

Yusuf F. Zakariya , H. K. Nilsen , Simon Goodchild & Kirsten Bjørkestøl

To cite this article: Yusuf F. Zakariya , H. K. Nilsen , Simon Goodchild & Kirsten Bjørkestøl (2020): Self-efficacy and approaches to learning mathematics among engineering students: empirical evidence for potential causal relations, International Journal of Mathematical Education in Science and Technology, DOI: 10.1080/0020739X.2020.1783006

To link to this article: https://doi.org/10.1080/0020739X.2020.1783006

Published online: 25 Jun 2020.

INTERNATIONAL JOURNAL OF MATHEMATICAL EDUCATION IN SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/0020739X.2020.1783006

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Self-efficacy and approaches to learning mathematics among engineering students: empirical evidence for potential causal relations

Yusuf F. Zakariya , H. K. Nilsen, Simon Goodchild and Kirsten Bjørkestøl

Department of Mathematical Sciences, University of Agder, Kristiansand, Norway

ABSTRACT
Theories of self-efficacy and approaches to learning are well-established in the psychology of learning. However, studies on relationships between the primary constructs on which these theories are developed are rarely reported in mathematics education research. Thus, the purpose of the current study is to provide empirical evidence for a potential causal relationship between perceived self-efficacy and approaches to learning. The present study adopts a cross-sectional survey research design that includes 195 engineering students enrolled on a first-year introductory calculus course. The data are collected using two well-developed and validated instruments with established high psychometric properties. Two hypotheses are formulated and tested using a structural equation modelling approach coupled with a weighted least square mean and variance adjusted estimator. The findings show that a high sense of perceived self-efficacy has a strong tendency to induce a deep approach to learning mathematics. In contrast, a low sense of perceived self-efficacy induces a surface approach to learning mathematics with a strong effect. This study represents a shift from the usual correlational studies that characterize quantitative research in mathematics education literature to causal relation research. Therein, causal assumptions are made and tested against the collected data, and some recommendations are made for future studies.

ARTICLE HISTORY
Received 25 March 2020

KEYWORDS
Self-efficacy; deep approach; surface approach; causal relation; empirical evidence

1. Introduction

Mathematics instruction that leads to satisfactory learning outcomes in terms of high performance as measured in examinations, understanding that supports future progression, engaged, motivated and enthusiastic students, has not been an easy task. Students, teachers, parents, researchers, policymakers, and other education stakeholders seek possible solutions to the global trend of poor performance in mathematics. The utility of mathematics transcends several educational levels, employment, and career opportunities, which explains why engineering students value the subject (Tossavainen et al., 2019). Mastery of introductory first-year mathematics courses is crucial to successful performance on core

CONTACT Yusuf F. Zakariya yusuf.zakariya@uia.no Department of Mathematical Sciences, University of Agder, Kristiansand, Norway

engineering courses at later years in the university. However, many first-year engineering students struggle with these courses, and their poor performance compels some of them to develop negative attitudes toward mathematics and change their career aspirations (Braathe & Solomon, 2015; Martínez-Sierra & García-González, 2016). Since students suffer most of the associated effects of poor performance in mathematics, a study that focuses on factors that emanate from the students is equally important. Several empirical studies have linked a variety of factors to poor performance in mathematics. These factors include but are not limited to mathematics anxiety (Dowker et al., 2016), attitudes toward mathematics (Dowker et al., 2019), academic motivation (Tossavainen et al., 2019), perceived self-efficacy (Williams & Williams, 2010), approaches to learning (Maciejewski & Merchant, 2016), conception of mathematics (Yang et al., 2019), prior mathematics knowledge (Zakariya, 2016), and self-concept (Pajares & Miller, 1994).

Two of these factors (perceived self-efficacy and approaches to learning) have received increased attention recently. The reason for this increased attention may lie in their satisfactory prediction of students' performance in mathematics (Loo & Choy, 2013; Maciejewski & Merchant, 2016; Williams & Williams, 2010). Perceived self-efficacy is linked to Albert Bandura's self-efficacy theory, which is grounded in the agentic social cognitive theory (Bandura, 1997). Perceived self-efficacy encapsulates 'beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments' (Bandura, 1997, p. 3). With a particular focus on engineering students, perceived self-efficacy has been defined as 'a person's belief that he or she can successfully navigate the engineering curriculum and eventually become a practicing engineer' (Jordan et al., 2010, p. 2). It is an important personal factor that facilitates improved students' performance in mathematics and boosts perseverance when undertaking difficult tasks (Bandura, 2012). Empirical studies have revealed that students with a high sense of perceived self-efficacy have low mathematics anxiety, high motivation to learn, positive attitudes toward mathematics, and increased interest in the subject (Bandura, 1997). Perceived self-efficacy has also been reported to predict students' performance in mathematics better than mathematics self-concept and prior knowledge of mathematics (Pajares & Miller, 1994). Efficacy beliefs have also been found to exert a more substantial direct effect on students' performance in a mathematics problem-solving activity than mental ability, mathematics anxiety, and high school mathematics content level (Pajares & Kranzler, 1995).

Students approach their learning of mathematics in different ways. However, these diverse ways of learning have been postulated by the approaches to learning theory to converge to two main approaches (Marton & Booth, 1997). Some engineering students learn mathematics with the motives of developing a deep understanding of its concepts (deep approach). In contrast, other students are extrinsically motivated to learn mathematics, such as satisfying the curriculum requirement, and thereby concentrate on crucial points (surface approach) to pass the course (Zakariya et al., 2020). Deep approaches to learning have generally been associated with an improved performance of first-year students on mathematics tasks more than surface approaches (Maciejewski & Merchant, 2016). However, there are some studies where a surface approach to learning mathematics has been reported to have a slightly higher positive correlation with performance than the deep approach to learning among engineering masters students (Svedin et al., 2013). Approaches to learning are strongly related to attitudes toward mathematics, conceptions of mathematics, and enjoyment of mathematics. Prior studies have shown a positive correlation between deep approach and attitudes toward mathematics and a negative correlation between surface approach and the latter (Alkhateeb & Hammoudi, 2006). The surface approach to learning predicts performance better than the enjoyment of mathematics, mathematics anxiety, motivation, the utility of mathematics, and gender (García et al., 2016).

Despite the success and satisfactory performance of both approaches to learning and perceived self-efficacy in predicting students' mathematics achievement, studies on causal relations between these constructs are rarely reported in the literature. Admittedly, some correlational studies are available which focus on science courses e.g. chemistry (Ardura & Galán, 2019), students enrolled on earth science programmes (Shen et al., 2016), and teachers in training (Phan, 2011). Thus, the purpose of the current study is to provide evidence for a potential causal relationship between perceived self-efficacy and approaches to learning among engineering students enrolled on a first-year calculus course. The present study is significant because if such a causal relation is revealed, then it is worth seeking interventions on one of the two constructs that can be designed to boost the other construct, which will, in turn, enhance students' performance. It is important to remark that the current study is not aimed at discovering an outright causal relation between the research constructs. Instead, causal assumptions are made therein to develop a model, and data are collected to test the causal model such that empirically-based arguments can be articulated to justify the plausibility of the model. As such, the main research question that this study attempts to address is: Does perceived self-efficacy influence the adoption of either deep or surface approach to learning mathematics among first-year engineering students?

2. Conceptual framework

A conceptual framework that can justify the relationship between approaches to learning mathematics and perceived self-efficacy among engineering students, rests on ideas from two psychological theories. Namely, approaches to learning theory and self-efficacy theory. The ontological and epistemological postulates of these theories and arguments that result in hypothesis formulations are presented in this section.

3. Student approaches to learning (SAL) theory

SAL theory can be linked to several studies of Marton and his colleagues on explorations and characterizations of approaches that university students adopt while reading some passages of prose and extracts of newspaper articles before being examined on their understanding of the presented materials (Marton & Säljö, 1976, 2005). Their qualitative analyses reveal diverse approaches to students' learning, which are highly motivated by prior experience, social factors, and the meanings that the students attached to learning (Marton & Booth, 1997). According to SAL theory, learning – a change in the experience of people about the world – forms a non-dualistic relationship between an individual and everything outside of it that is neither individually constructed nor environmentally imposed (Marton & Booth, 1997). Thus, it can be argued that students' approaches to learning vary because of the feedback relationship between students' motivation to learn, intentions, and learning context. However, these various students' approaches to learning can be generally classified into deep and surface approaches (Marton & Säljö, 2005).

Biggs (2012) describes deep approaches to learning as 'activities that are appropriate to handling the task so that an appropriate outcome is achieved' while surface approaches to learning, on the other hand, encapsulate 'activities of an inappropriately low cognitive level, which yields fragmented outcomes that do not convey the meaning of the encounter' (p.42). As such, considering the nature of engineering programmes in which students are being trained to solve practical problems, it is expected, if not required, that students adopt approaches that will facilitate the development of high cognitive skills required to solve these problems. Furthermore, approaches to learning according to SAL tradition (Marton & Booth, 1997) are predictable from students' learning conception – 'a qualitatively distinct manner in which the subjects were found to voice the way they thought about learning' (p.36), motives, intents, and the learning situations. For instance, engineering students who conceive calculus tasks as something useful and which proper understanding of it is necessary for intellectual development are likely to adopt deep approaches to learning the course. On the other hand, students who conceive calculus tasks as a mere requirement to move to the next level of study are likely to adopt surface approaches to learning the course. Thus, a deep approach to learning is intrinsically motivated, while a surface approach to learning is extrinsically motivated (Hounsell, 2005; Marton & Säljö, 2005).

It is important to remark that learning situations in the context of mathematics learning also include the nature of mathematics tasks. Such that the approaches students adopt to learning the subject are highly influenced by the nature of the tasks. Maciejewski and Merchant (2016), in an empirical cross-sectional study, show that there is a strong correlation between a deep approach to learning and students' first-year grades on mathematics tasks while a surface approach to learning has no significant correlation. However, for year-two, year-three, and year-four students, there is a strong negative correlation between the surface approach to learning and students' grades in which a deep approach shows no significant correlation. These discrepancies and inconsistencies in strength and direction of correlation coefficients between approaches to learning mathematics and students' grades are argued, using Bloom's taxonomy, to stem from the different nature of mathematics tasks at the different years of study (Maciejewski & Merchant, 2016). As such, considering its task specificity, approaches to learning mathematics are best investigated by focusing on a set of students who are following a common mathematics course.

4. Self-efficacy theory

Perceived self-efficacy is an essential component of the agentic social cognitive theory that describes behavioural changes of an individual as continuously being modified and regulated through a feedback interaction with social factors (Bandura, 2001). Unlike the traditional social cognitivism, it is argued that both social structure and personal agency 'function interdependently rather than as disembodied entities' (Bandura, 2012, p. 15). Thus, a rejection of an ontological position of dualism between social structure and personal agency. As such, agentic social cognitive theory relies on an epistemological proposition called 'reciprocal determinism' introduced by Bandura (1986, 2012). Reciprocal determinism describes human functioning as a triadic feedback causal model between personal, environmental, and behavioural factors. Therefore, it can be argued that perceived self-efficacy of engineering students on mathematics tasks is not a fixed construct since it an integral part of the personal factors that are embedded in the reciprocal deterministic model. Instead, it is causally affected by changes in the model. Borgonovi and Pokropek (2019) elaborate more on this concept when they write 'reciprocal determinism describes the sets of relationships underlying the interactions between: (a) individuals' exposure to mathematics tasks, (b) mathematics self-efficacy beliefs, and (c) mathematics ability' (p. 269).

Perceived self-efficacy contributes significantly to regulating affective, cognitive, decisional, and motivational processes of human functioning (Bandura, 2001, 2002). It is an essential construct in the learning process as it serves as a stimulus for students not to give up on difficult learning situations such that desired outcomes are achieved. It makes the individual's involvement very active and boosts morale to see to the attainment of a desirable outcome (Bandura, 1997, 2012). Since self-efficacy beliefs regulate some decisional processes of a learner, it can be argued that there is a causal relationship between perceived self-efficacy and approaches to learning mathematics. This is because students' approaches to learning a content are crucial components of their decisional processes (Biggs, 1993). Another proxy construct through which perceived self-efficacy can be causally linked with approaches to learning is students' motivation. Intrinsic motivation has been shown to induce a deep approach to learning while extrinsic motivation to induce a surface approach to learning (Marton & Booth, 1997). As such, it is expected that perceived self-efficacy is causally related to deep and surface approaches through motivation as an intervening construct since self-efficacy beliefs regulate motivational processes (Bandura, 1997).

To substantiate the argument on the causal relationship between approaches to learning mathematics and perceived self-efficacy, one could also turn to some findings that have been reported in other fields. For instance, Diseth (2011), in a study involving 177 first-year undergraduate students following a psychology course used a causal model to expose a negative relation between self-efficacy and surface approaches to learning, and an indirect positive relationship between self-efficacy and deep approaches to learning. The study by Shen et al. (2016) also reports a strong positive relationship between the deep approach to learning earth sciences and perceived self-efficacy. After an extensive search of the literature, the only quantitative study the authors could find on approaches to learning mathematics, and perceived self-efficacy is a correlational study by Zakariya et al. (2019). Therein, deep approaches to learning mathematics are found to have a positive correlation with perceived self-efficacy on calculus tasks, and a negative correlation is found between the latter and surface approaches to learning. Thus, based on the aforementioned discussion, the following hypotheses are formulated.

> *Hypothesis one*: There is a positive causal effect of perceived self-efficacy on deep approaches to learning a first-year introductory calculus course among engineering students.
>
> *Hypothesis two*: There is a negative causal effect of perceived self-efficacy on surface approaches to learning a first-year introductory calculus course among engineering students.

5. Methodology

5.1. Participants

The focus of the current study, using a cross-sectional survey research design, is on first-year engineering students at a leading Norwegian university. Even though they are enrolled on different engineering programmes, they followed a common introductory first-semester calculus course at the university. A total of 195 (47 females) students who voluntarily gave their consent took part in the study. The sample corresponds to about 65% of the total population of first-year engineering students who were invited to participate in the study. This response rate is considered high in the literature (Babbie, 1990).

5.2. Materials

Two well-developed survey instruments were used for collecting the research data. The first instrument was a Norwegian language adaptation of the revised two-factor study process questionnaire (R-SPQ-2F). This instrument was initially conceptualized and operationalized based on SAL theory to measure students' approaches to learning by Biggs et al. (2001) and was adapted to mathematics learning context among Norwegian first-year engineering students by Zakariya et al. (2020). The Norwegian adaption of the R-SPQ-2F is a 19-item questionnaire that measures two dimensions (deep and surface) of approaches to learning mathematics on a five-point Likert scale from (1) *never or only rarely true of me*, through (3) *it is true of me about half the time,* to (5) *it is always or almost always true of me.* The deep approach subscale has ten items with a reliability coefficient of .81, and the surface approach subscale has nine items with a reliability coefficient of .72 (Zakariya et al., 2020). The construct validity of the Norwegian adaption of R-SPQ-2F has been studied involving several comparisons of competing models using confirmatory factor analyses (Zakariya, 2019). Despite the availability of other measuring instruments of students' approaches to learning, such as the approaches and study skills inventory for students (ASSIST), R-SPQ-2F was adopted in the current study for a few reasons. First, it has been validated and available in Norwegian, which is the main language of instruction in the university undergraduate programmes. Second, it is concise with only 19 items, unlike the ASSIST, with 52 items and has strong psychometric characteristics. Third, given that approaches to learning are context-specific, an adapted R-SPQ-2F into mathematics context is likely to possess a high predictive power.

The second instrument used for collecting data in the current study was a calculus self-efficacy inventory (CSEI) developed by (Zakariya et al., 2019). The CSEI is a 13-item instrument developed based on guidelines for constructing perceived self-efficacy scales as explicated by the Bandura's self-efficacy theory (Bandura, 2006). The inventory contains calculus final exam-like questions in which the students are required to rate how much confidence they have in solving the questions correctly on a scale from 0 to 100. It was found to have high construct validity with unidimensionality of its items, high discriminant validity, and a high reliability index of .90 (Zakariya et al., 2019). The CSEI was adopted in the current study not only for its strong psychometric properties and because its theoretical foundation suits our conceptual framework but also for its specificity in measuring student perceived self-efficacy on calculus tasks. The Norwegian adaption of R-SPQ-2F and the CSEI were embedded in an on-line survey tool and administered to the students via their email addresses. The data collection exercise took about two weeks, and the collected data were screened for outliers and missing values. There was no case of outliers, and few data were missing at random, which were less than 1% of the total data collected and, as such, do not pose any challenge to the analyses.

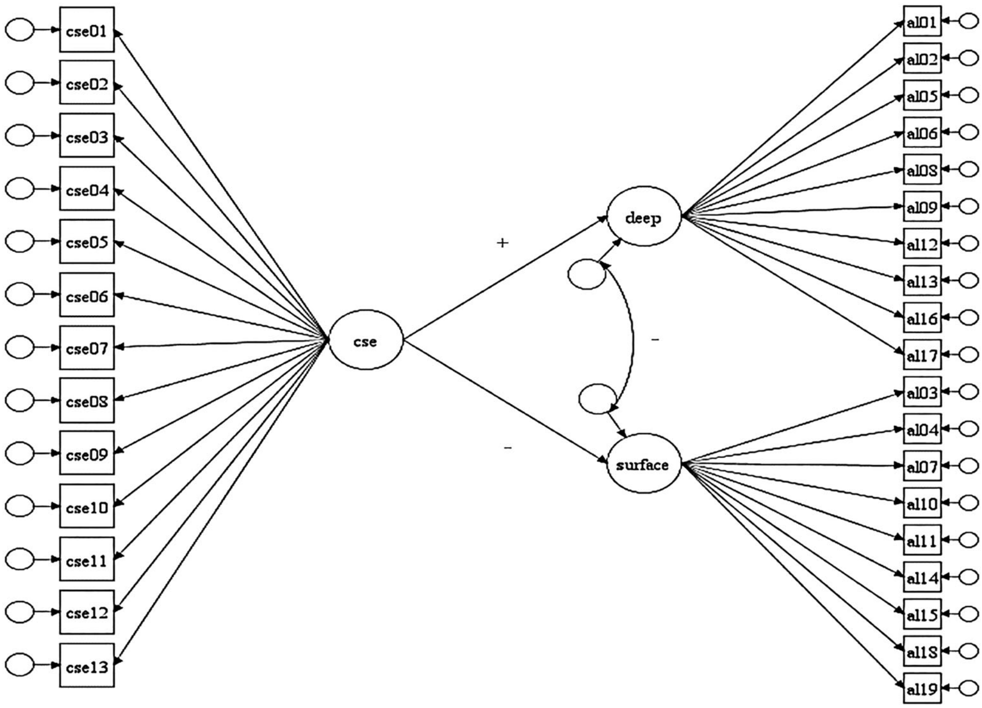


Figure 1. A hypothesized causal model of the relations between perceived calculus self-efficacy and approaches to learning mathematics.

5.3. Data analysis

The collected data were analysed using a structural equation modelling (SEM) approach. This involved testing the plausibility of the hypothesized causal relations between perceived self-efficacy and approaches to learning mathematics, as shown in Figure 1. As succinctly put by Bollen and Pearl (2013), 'SEM is an inference engine that takes in two inputs, qualitative causal assumptions, and empirical data, and derives two logical consequences of these inputs: quantitative causal conclusions and statistical measures of fit for the testable implications of the assumptions' (p.309). It is thus argued that the use of SEM in exposing the causal relationship between the current research constructs is justified, and an alternative statistical model that can do a satisfactory job in evaluating these causal claims is unlikely (Bullock et al., 1994). All the model parameters such as factor loadings, effect weights, residuals, and intercepts were evaluated using the weighted least square mean and variance adjusted (WLSMV) estimator with theta parameterization. WLSMV was used because of its satisfactory high performance in the analysis of categorical data such as the ones obtained using the Likert scale (Suh, 2015).

Figure 1 presents a graphical representation of the hypotheses one and two of the current study. The big oval shape with a label 'cse' represents the latent variable of the students' perceived calculus self-efficacy (henceforth refers to as self-efficacy) as measured by its 13 observed variables (rectangles with labels 'cse01' to 'cse13') accompanied by the small oval shapes with small arrows pointing to each rectangle indicating the associated errors in predicting each observed variable. In a similar manner, the big oval shapes with the labels 'deep' and 'surface' represent the latent variables of deep and surface approaches to learning mathematics, respectively, each of which is measured by its respective number of observed variables ('al01'-'al19'). The single-headed arrow between 'cse' and 'deep' and the one between 'cse' and 'surface' both indicate the hypothesized causal relations between these latent constructs with their respective signs as postulated in the hypotheses one and two. The double-headed arrow with a negative sign is an expected negative correction between deep and surface approaches. This is because a student with a high score on surface approach items of the R-SPQ-2F is expected to have a low score on the deep approach items.

The causal model presented in Figure 1 carries with it a few assumptions that are subject to testability. Prominent assumptions are represented by causal arrows from the self-efficacy to the two dimensions of approaches to learning. The observed variables are also assumed to relate to their respective latent constructs in a linearly causal manner. The errors of the observed variables are assumed to be uncorrelated with each other and with any of the latent constructs. It is also assumed that none of the observed variables exhibits a cross-loading i.e. each observed variable is assumed to expose only one latent construct. These qualitative assumptions are the elements that make a whole of the causal model presented in Figure 1 on which data are collected, analysed, and their plausibility is ascertained using some goodness of fit (GOF) indices. The following GOF indices are used to judge an acceptable fit: Tucker-Lewis index (TLI) and comparative fit index (CFI) with values close to or greater than .90 (Bentler, 1990), standardized root mean square residual (SRMR) with a value less than .80 (Hu & Bentler, 1999), and root mean square error of approximation (RMSEA) with a value less than or equal to .10 (MacCallum et al., 1996). Chi-square statistics are reported, and its ratio to the degree of freedom of less or equal to 3 (Brown, 2015) is used to assess a model fit. Also, chi-square statistics are also used to compare competing models using a difference test.

6. Results

6.1. Measurement model evaluations

The evaluation of the hypothesized causal model presented in Figure 1 proceeds in two steps. The first step concerns fitting a separate measurement model for both CSEI and R-SPQ-2F. This step is a preliminary step to the structural equation modelling of the relations between the research constructs. The ensuing results are presented in Table 1 with Model 1 for the CSEI measurement model and Model 2 for the R-SPQ-2F measurement model. Table 1 also presents improved results for both Model 1 and Model 2.

The results presented in Table 1 reinforce a rejection of Model 1. Consequently, a rejection of some assumptions associated with this model. This is evident with a high ratio of chi-square value to *df* (higher than 3), a low TLI value (lower than .90), and a high RMSEA value (higher than .10). Thus, the results show some inconsistency between the data collected and the hypothesized model. As such, Model 1 was improved upon by adding two error covariances between 'CSE09' and 'CSE11' as well as between 'CSE12' and 'CSE13', the results of which are presented in Table 1 (Improved Model 1). There is a significant improvement in Model 1 after modifying it, as shown in Table 1 with a reduced ratio of

Table 1. Selected GOF indices for the evaluations of CSEI and R-SPQ-2F measurement models.

	CSEI		R-SPQ-2F	
Fit statistics	Model 1	Improved Model 1	Model 2	Improved Model 2
Chi-square value (χ^2)	250.18	184.92	461.92	300.15
Degree of freedom (*df*)	65	63	151	148
χ^2 / *df*	3.85	2.94	3.06	2.03
TLI	.88	.92	.77	.89
CFI	.90	.94	.80	.90
SRMR	.07	.06	.09	.07
RMSEA	.12	.10	.10	.07

chi-square value to *df*, improved TLI and CFI, and reduced SRMR and RSMEA. The chi-square difference test exposes this improvement better as it returns a significant difference in chi-square values ($\chi^2(2) = 65.26, p < .001$) between Model 1 and Improved Model 1. These are all suggestive of the plausibility of the Improved Model 1.

Similarly, the results presented in Table 1 (Model 2) also show that the hypothesized measurement model for the R-SPQ-2F should be rejected. This is evident with a high ratio of chi-square value to *df* (higher than 3), a low TLI and CFI values (far lower than .90), and a high SRMR value (higher than .08). As such, Model 2 was improved upon by allowing 'al10' and 'al19' to cross-load on the deep approach to learning in addition to the surface approach to learning they were initially hypothesized. Also, an error covariance between 'al15' and 'al18' was included to achieve the model results presented in Table 1 (Improved Model 2). As it can be read from Table 1 (Improved Model 2), all the GOF indices are within the cutoff criteria coupled with a significant chi-square difference test ($\chi^2(3) = 161.77, p < .001$) which affirm the plausibility of the improved version of Model 2. The model modifications that have been carried out in this section are all guided by modification indices of the respective output during the analyses, and its conceptual implications are presented in the next section.

6.2. Structural model evaluations

After the validation of the measurement models, we proceed to the second step of the analyses, which concerns investigating the causal relations between self-efficacy and approaches to learning mathematics. The ensuing results of the selected fit statistics show the ratio of chi-square value to *df* to be 1.59, a TLI value to be .90, a CFI value to be .91, an SRMR value to be .07, and an RMSEA value to be .06 which are suggestive of an acceptable fit of the model. The causal estimates, as well as the associated standardized model parameters such as factor loadings, factor variance, and error covariance, are presented in Figure 2.

The results presented in Figure 2 show all the significant standardized factor loadings and error covariances. Figure 2 shows that there is a significant positive causal effect of the self-efficacy on deep approaches to learning a first-year introductory calculus ($\beta = .54, p < .001$) with a medium significant effect size of .29, and a significant negative causal effect of self-efficacy on surface approaches to learning a first-year introductory calculus ($\beta = -.47, p < .001$) with a medium significant effect size of .22. The results confirm the hypotheses one and two, respectively. These results can be interpreted to mean, given a unit metric increase in self-efficacy (e.g. $cse + 1$) there is a corresponding effect of a .54

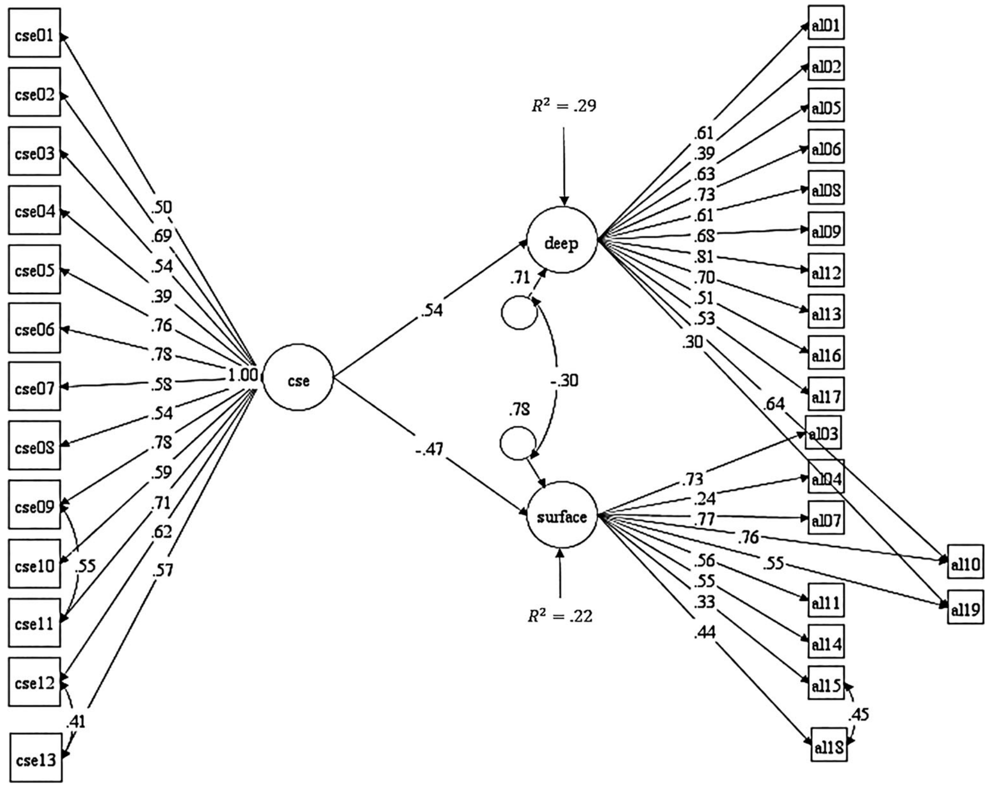


Figure 2. A validated causal model of the relations between self-efficacy and approaches to learning mathematics.

times a unit metric increase on deep approaches to learning, and a corresponding effect of a .47 times a unit metric decrease in surface approaches to learning among engineering students. Further, the respective effect sizes show that self-efficacy accounts for 29% of the variability in deep approaches to learning and 22% variability in the surface approaches to learning. At this juncture, it is important to remark that the results presented in Figure 2 are valid up to the group level, and there could be some discrepancy when it comes to each individual student that took part in the study. Further discussion on these results is presented in the next section.

7. Discussion, limitations, and conclusion

7.1. Discussion

The current study attempts to provide evidence for possible causal relations between self-efficacy and approaches to learning an introductory calculus course among first-year engineering students. In order to achieve this, both measurement and structural evaluations of models based on data collected using CSEI and R-SPQ-2F are reported in the current study. We improved on the CSEI measurement model by adding two error covariances. The first error covariance was between 'CSE09' and 'CSE11'. These two items measure students' confidence in solving two different indefinite integral tasks. As such, an error covariance between these items could account for any error source from the common topic from which these two items have been drawn. The second error covariance was between 'CSE12' and 'CSE13', which can also be justified from a common topic (applications of integral calculus) from which the two items have been drawn. The addition of these two error covariances negates the lack of it that was initially assumed in the CSEI measurement model. Thus, it demonstrates how the plausibility of model assumptions can be tested in the SEM framework, which is contrary to those who think SEM assumptions are never tested (e.g. Freedman, 1995). Further, the addition of these error covariances also shows a marked difference between SEM framework and regression models because, in the latter, errors are always assumed to be orthogonal i.e. uncorrelated with each other (Bollen & Pearl, 2013).

The measurement model of the R-SPQ-2F was also improved upon by allowing items al10 ('I find I can get by in most assessments by memorizing key sections rather than trying to understand them') and al19 ('I find the best way to pass examinations is to try to remember answers to likely questions') that are initially on the surface approach dimension to cross-load on the deep approach dimension (Zakariya et al., 2020). A common aspect of these two items revolves around the memorization of key concepts. The finding of the current study suggests that the memorization technique is not peculiar to surface approach learners. Rather, students that adopt a deep approach to learning mathematics may also use the memorization technique. This finding, on the one hand, corroborates a report of widespread use of the memorization technique found among high achieving Asian students as a means of understanding (Kember, 1996). On the other hand, it suggests Entwistle (1997) could be right when he wrote: 'memorization is a necessary precursor to understanding, and for other purposes it is a way of reinforcing understanding'. Thus, a deep approach learner can as well use memorization techniques strategically to recall definitions of concepts, theorems, and procedures of carrying out some special differentiation or integration in a first-year calculus course.

Another model improvement of the R-SPQ-2F is the addition of an error covariance between al15 ('I believe that lecturers shouldn't expect students to spend significant amounts of time studying material everyone knows won't be examined') and al18 ('I see no point in learning material which is not likely to be in the examination') which are both measuring surface approaches to learning (Zakariya et al., 2020). This error covariance seems to be conceptually justified as both items share a common latent factor and emphasise skipping materials that are not going to be on students' examination questions. Moreover, this finding also fits into the body research that has advocated the inclusion of error covariances between some other items of surface approach dimension of the R-SPQ-2F (e.g. Önder & Besoluk, 2010).

Of prime importance in the current study is the established potential causal relation between self-efficacy and approaches to learning mathematics. It was found that self-efficacy has a positive effect on the deep approach to learning and a negative effect on the surface approach to learning. This seems to be the first time such a finding is being reported on mathematics learning of engineering students. However, our findings, on the one hand, do support a negative relation between self-efficacy and approaches to learning among psychology students reported by Diseth (2011). On the other hand, our findings establish a reverse relationship as compared to the report by Ardura and Galán (2019) on self-efficacy and approaches to learning Physics and Chemistry among secondary school students. Therein, Ardura and Galán (2019) proposed, tested, and found small effects (from − .12 to .25) between the dimensions of approaches to learning on self-efficacy. The reported potential causal effects between self-efficacy and approaches to learning mathematics in the current study are far away higher than the reverse effects by Ardura and Galán (2019). These suggest that our model establishes a better and more appropriate causal direction of the relationship between these constructs. Further, the estimates of the potential causal effects between self-efficacy and approaches to learning in the current study are higher than the correlation coefficients between these constructs in earth sciences (Shen et al., 2016) and in mathematics (Zakariya et al., 2019).

Even though, the percentages of explained variance in deep approaches to learning (29%) and surface approaches to learning (22%) that are accounted by the self-efficacy seem low in the current study they are substantially higher than reported values in the literature (e.g. Diseth, 2011). These percentages of explained variance are reflections of the fact that there are other factors e.g. motivation, nature of mathematics tasks, etc., that influence approaches to learning mathematics, which our proposed model does not account for. Admittedly, we do not seek to propose a model that explains every relation between self-efficacy and approaches to learning mathematics. Instead, the current study has attempted to provide evidence for a potential causal relationship between these constructs. The findings of the current study, therefore, will serve as justifications of designing self-efficacy interventions by university lecturers, engineering course coordinators, and other stakeholders who are directly involved in the teaching of mathematics to engineering students as proxies to induce desired learning approaches in their students.

7.2. Limitations

Despite the promising strength of the current study in providing evidence on the causal relation between the two important student-source factors, some limitations can be acknowledged. First, the current study is confined within the two research constructs without relating the ensuing effects to students' performance in the introductory calculus course. The authors acknowledge that it would have been more interesting to see how these effects translate either directly or indirectly from self-efficacy through approaches to learning to students' grades in the course. However, students' grades are not included in the model because of the unavailability of these grades to the researchers at the time of the study. A future study will be conducted with this intention. Second, given that the scope of the current study is limited to first-year engineering students at one university, its findings might be limited to this student population. Perhaps, the inclusion of students from year two, year three and year four or students following other programmes in the study and other institutions would have increased the generalization power of its findings. Third, and closely related to the second limitation is the lack of cross-validation of the established structural model of the relation self-efficacy and approaches to learning. To this end, we recommended replicated evaluations of this model in independent samples and across different student populations. Lastly, the authors declare that the model proposed and evaluated in the current study is neither an absolute model nor a simplification of reality between the research constructs. Rather, attempts have only been made to understanding the complex relationship between these constructs from a theoretical and empirical perspective. Future in-depth analyses are recommended using case studies, longitudinal design studies, and experimental design studies.

8. Conclusion

The current study was motivated by the dearth of studies on relationships between two well-established psychological theories that concern students' learning of mathematics in higher education. Therein, empirical evidence is provided for a potential causal relation between self-efficacy and approaches to learning a first-year calculus course among engineering students. A high sense of self-efficacy is found to induce the adoption of a deep approach to learning, while low sense self-efficacy induces an adoption of a surface approach to learning. We claim this to be an original contribution to the literature when it comes to the learning of mathematics among engineering students. As such, more studies are recommended using diverse methodological approaches and designs to understand further the relations between self-efficacy and approaches to learning mathematics.

Disclosure statement

The authors declare that there is no potential conflict of interest

ORCID

Yusuf F. Zakariya http://orcid.org/0000-0002-5266-8227

References

Alkhateeb, H. M., & Hammoudi, L. (2006). Attitudes toward and approaches to learning first-year university mathematics. *Perceptual and Motor Skills*, *103*(5), 115–120. https://doi.org/10.2466/PMS.103.5.115-120

Ardura, D., & Galán, A. (2019). The interplay of learning approaches and self-efficacy in secondary school students' academic achievement in science. *International Journal of Science Education*, *41*(13), 1723–1743. https://doi.org/10.1080/09500693.2019.1638981

Babbie, E. R. (1990). *Survey research methods*. Wadsworth.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology: An International Review*, *51*(2), 269–290. https://doi.org/10.1111/1464-0597.00092

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-Efficacy beliefs of Adolescents* (Vol. 5., pp. 307–337). Information Age Publishing.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, *38*(1), 9–44. https://doi.org/10.1177/0149206311410606

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Biggs, J. B. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Educational Research Journal*, *63*, 3–19. https://doi.org/10.1111/j.2044-8279.1993.tb01038.x

Biggs, J. B. (2012). What the student does: Teaching for enhanced learning. *Higher Education Research & Development*, *31*(1), 39–55. https://doi.org/10.1080/07294360.2012.642839

Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, *71*(1), 133–149. https://doi.org/10.1348/000709901158433
Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation. In S. L. Morgan (Ed.), *Handbooks of Sociology and social research* (pp. 301–328). Springer.
Borgonovi, F., & Pokropek, A. (2019). Seeing is believing: Task-exposure specificity and the development of mathematics self-efficacy evaluations. *Journal of Educational Psychology*, *111*(2), 268–283. https://doi.org/10.1037/edu0000280
Braathe, H. J., & Solomon, Y. (2015). Choosing mathematics: The narrative of the self as a site of agency. *Educational Studies in Mathematics*, *89*(2), 151–166. https://doi.org/10.1007/s10649-014-9585-8
Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.
Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling: A Multidisciplinary Journal*, *1*(3), 253–267. https://doi.org/10.1080/10705519409539977
Diseth, Å. (2011). Self-efficacy, goal orientations and learning strategies as mediators between preceding and subsequent academic achievement. *Learning and Individual Differences*, *21*(2), 191–195. https://doi.org/10.1016/j.lindif.2011.01.003
Dowker, A., Cheriton, O., Horton, R., & Mark, W. (2019). Relationships between attitudes and performance in young children's mathematics. *Educational Studies in Mathematics*, *100*(3), 211–230. https://doi.org/10.1007/s10649-019-9880-5
Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, 508. https://doi.org/10.3389/fpsyg.2016.00508
Entwistle, N. (1997). Reconstituting approaches to learning: A response to Webb. *Higher Education*, *33*(2), 213–218. https://doi.org/10.1023/A:1002930608372
Freedman, D. A. (1995). Causal diagrams for empirical research: Discussion of 'causal diagrams for empirical research' by J. Pearl. *Biometrika*, *82*(4), 692–693. https://doi.org/10.1093/biomet/82.4.692
García, T., Rodríguez, C., Betts, L., Areces, D., & González-Castro, P. (2016). How affective-motivational variables and approaches to learning predict mathematics achievement in upper elementary levels. *Learning and Individual Differences*, *49*, 25–31. https://doi.org/10.1016/j.lindif.2016.05.021
Hounsell, D. (2005). Understanding teaching and teaching for understanding. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 238–257). University of Edinburgh, Centre for Teaching, Learning and Assessment.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Jordan, K. L., Sorby, S., Amato-Henderson, S., & Donahue, T. H. (2010). *Engineering self-efficacy of women engineering students at urban vs. rural universities* [Paper presented]. The 41st ASEE/IEEE Frontiers in Education Conference, Rapid City, South Dakota, USA.
Kember, D. (1996). The intention to both memorise and understand: Another approach to learning? *Higher Education*, *31*(3), 341–354. https://doi.org/10.1007/BF00128436
Loo, C. W., & Choy, J. L. F. (2013). Sources of self-efficacy Influencing academic performance of engineering students. *American Journal of Educational Research*, *1*(3), 86–92. https://doi.org/10.12691/education-1-3-4
MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130
Maciejewski, W., & Merchant, S. (2016). Mathematical tasks, study approaches, and course grades in undergraduate mathematics: A year-by-year analysis. *International Journal of Mathematical Education in Science and Technology*, *47*(3), 373–387. https://doi.org/10.1080/0020739X.2015.1072881

Martínez-Sierra, G., & García-González, M. D. S. (2016). Undergraduate mathematics students' emotional experiences in Linear Algebra courses. *Educational Studies in Mathematics*, *91*(1), 87–106. https://doi.org/10.1007/s10649-015-9634-y

Marton, F., & Booth, S. A. (1997). *Learning and awareness*. Lawrence Erlbaum.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning I: Outcome and process. *British Journal of Educational Psychology*, *46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Marton, F., & Säljö, R. (2005). Approaches to learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 39–58). University of Edinburgh, Centre for Teaching, Learning and Assessment.

Önder, I., & Besoluk, S. (2010). Adaptation of revised two factor study process questionnaire (R-SPQ-2F) to Turkish. *Education and Science*, *35*(157), 55–67.

Pajares, F., & Kranzler, J. H. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, *20*(4), 426–443. https://doi.org/10.1006/ceps.1995.1029

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, *86*(2), 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Phan, H. P. (2011). Interrelations between self-efficacy and learning approaches: A developmental approach. *Educational Psychology*, *31*(2), 225–246. https://doi.org/10.1080/01443410.2010.545050

Shen, K.-M., Lee, M.-H., Tsai, C.-C., & Chang, C.-Y. (2016). Undergraduate students' earth science learning: Relationships among conceptions, approaches, and learning self-efficacy in Taiwan. *International Journal of Science Education*, *38*(9), 1527–1547. https://doi.org/10.1080/09500693.2016.1198060

Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(4), 568–580. https://doi.org/10.1080/10705511.2014.937669

Svedin, M., Balter, O., Scheja, M., & Pettersson, K. (2013). *A surface approach to learning rewards first-year engineering students*. [Paper presented]. The 2013 Learning and Teaching in Computing and Engineering.

Tossavainen, T., Rensaa, R. J., & Johansson, M. (2019). Swedish first-year engineering students' views of mathematics, self-efficacy and motivation and their effect on task performance. *International Journal of Mathematical Education in Science and Technology*, 1–16. https://doi.org/10.1080/0020739X.2019.1656827

Williams, T., & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology*, *102*(2), 453–466. https://doi.org/10.1037/a0017271

Yang, X., Leung, F., & Zhang, S. (2019). Junior secondary school students' conceptions of and approaches to learning mathematics and their relationships in mainland China. *Sustainability*, *11*(9), 2476. https://doi.org/10.3390/su11092476

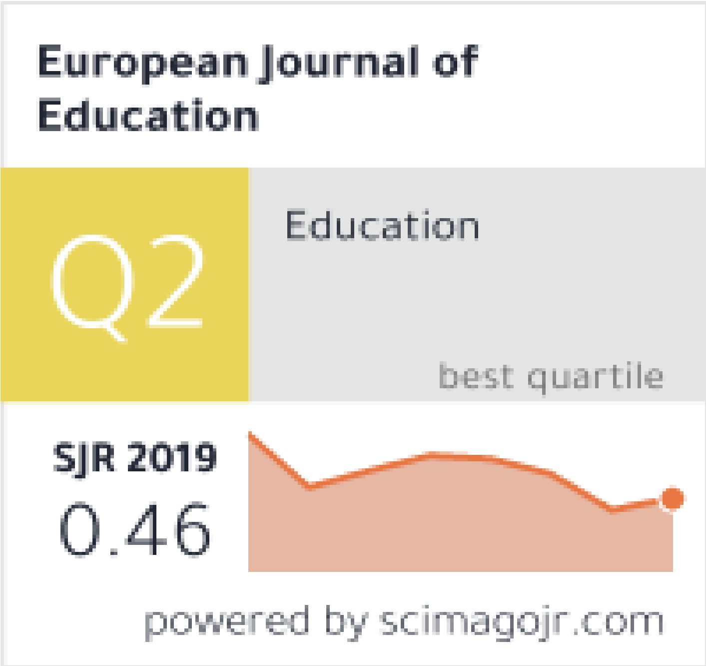
Zakariya, Y. F. (2016). Investigating validity of Math 105 as prerequisite to Math 201 among undergraduate students, Nigeria. *International Journal of Instruction*, *9*(1), 107–118. https://doi.org/10.12973/iji.2016.919a

Zakariya, Y. F. (2019). Study approaches in higher education mathematics: Investigating the statistical behaviour of an instrument translated into Norwegian. *Education Sciences*, *9*(3), 191. https://doi.org/10.3390/educsci9030191

Zakariya, Y. F., Bjørkestøl, K., Nilsen, H. K., Goodchild, S., & Lorås, M. (2020). University students' learning approaches: An adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Education Evaluation*, 100816. https://doi.org/10.1016/j.stueduc.2019.100816

Zakariya, Y. F., Goodchild, S., Bjørkestøl, K., & Nilsen, H. K. (2019). Calculus self-efficacy inventory: Its development and relationship with approaches to learning. *Education Sciences*, *9*(3), 170. https://doi.org/10.3390/educsci9030170

Paper VII



Zakariya, Y. F., Nilsen, H. K., Bjørkestøl, K., & Goodchild, S. (forthcoming). Effects of prior mathematics knowledge and approaches to learning on performance in mathematics among first-year engineering students. *European Journal of Education*, under review.

European Journal of
EDUCATION
RESEARCH, DEVELOPMENT AND POLICY

Effects of prior mathematics knowledge, and approaches to learning on performance in mathematics among first-year engineering students

Journal:	*European Journal of Education*
Manuscript ID	Draft
Wiley - Manuscript type:	Original Article
Keywords:	approaches to learning, higher education, mediation analysis, prior mathematics knowledge, structural equation modelling

Abstract:	This study is aimed at unravelling the specific effects of prior mathematics knowledge and approaches to learning on performance in mathematics among a convenient sample of 188 engineering students. The design is cross-sectional, and data were analysed with structural equation modelling. We found a positive effect of prior mathematics knowledge on performance and a negative effect of the former on surface approaches to learning. The effect of surface approaches to learning on performance is significant, negative, and surface approaches to learning mediate the effect of prior mathematics knowledge on performance. There are no substantial relations between prior mathematics knowledge, deep approaches to learning, and performance. Even though students who adopt surface approaches performed poorly, we found no evidence to claim that students who adopt deep approaches perform better in the course. By implication, our findings underscore the importance of discouraging engineering students from capitalising on surface approaches to learning mathematics.

Effects of prior mathematics knowledge, and approaches to learning on performance in mathematics among first-year engineering students

Abstract

This study is aimed at unravelling the specific effects of prior mathematics knowledge and approaches to learning on performance in mathematics among a convenient sample of 188 engineering students. The design is cross-sectional, and data were analysed with structural equation modelling. We found a positive effect of prior mathematics knowledge on performance and a negative effect of the former on surface approaches to learning. The effect of surface approaches to learning on performance is significant, negative, and surface approaches to learning mediate the effect of prior mathematics knowledge on performance. There are no substantial relations between prior mathematics knowledge, deep approaches to learning, and performance. Even though students who adopt surface approaches performed poorly, we found no evidence to claim that students who adopt deep approaches perform better in the course. By implication, our findings underscore the importance of discouraging engineering students from capitalising on surface approaches to learning mathematics.

Keywords: approaches to learning, higher education, mediation analysis, prior mathematics knowledge, structural equation modelling

INTRODUCTION

Students approach learning in different ways depending on the learning context, motivation to learn, intentions, self-efficacy, and nature of course assessment. Some students approach learning with concentration and seek a proper understanding of content materials, while others give the impression that for them learning is an obstacle or a requirement to proceed to the next stage of study or employment and, as such, rely heavily on memorising facts and procedures. The former students are said to adopt deep approaches to learning while the latter students are said to adopt surface approaches to learning (Entwistle, 1997; Zakariya, Bjørkestøl, Nilsen, Goodchild, & Lorås, 2020). Some researchers have identified a third approach (strategic) in which students effectively organise resources and manage their time efficiently with the intention of getting high grades in their examinations (Tait, Entwistle, & McCune, 1998). However, Biggs, Kember, and Leung (2001) have demonstrated how a strategic approach can be subsumed in the deep and surface approaches to learning. It has been theoretically argued that approaches to learning, whether as adopted processes before a learning activity that directly affect learning outcomes (Marton & Säljö, 2005) or as students' tendencies to adopt a process/strategy in a learning situation as noted by Biggs (1993), have a decisive role in the effectiveness of teaching and learning. There is an accumulation of evidence that suggests a relationship between approaches to learning and conceptions of mathematics (Mji, 2000), attitudes toward mathematics and mathematics anxiety (Alkhateeb & Hammoudi, 2006; Rozgonjuk, Kraav, Mikkor, Orav-Puurand, & Täht, 2020), and mathematics self-efficacy (Zakariya, Nilsen, Goodchild, & Bjørkestøl, 2020).

Generally, approaches to learning have been found to relate with performance and shown to be better predictors of performance than mathematics anxiety, gender, enjoyment of mathematics, the utility of mathematics, and intrinsic motivation (García, Rodríguez, Betts, Areces, & González-Castro, 2016). However, results of studies on specific effects of either deep or surface approaches to learning on students' performance have been inconclusive. Some researchers found that a deep approach to learning predicts performance, while surface approach does not do so. For instance, Guo, Yang, and Shi (2017) found, in a study involving 74687 university students drawn from 39 different universities in China on a variety of courses (42% of whom enrolled on science and engineering programmes), that there is a positive effect of a deep approach to learning on students' learning outcomes. Moreover, a negative non-significant impact of a surface approach to learning was found on students' learning outcomes. One may challenge the validity of the findings by Guo et al. (2017) because a self-report measure of learning outcomes was used in the study. However, there have been similar findings whereby the deep approach to learning predicts students' performance and the surface approach to learning does not predict students' performance. Therein, performance is measured by either a semester final examination scores (Cano, Martin, Ginns, & Berbén, 2018) or by course grades (Maciejewski & Merchant, 2016).

In contrast, a few researchers found that the surface approach to learning predicts academic performance, while the deep approach does not do so (Diseth, 2007; Trigwell, Ellis, & Han, 2012). The contribution of approaches to learning, among other factors, to 442 students' end of the semester performance in a first-year psychology course was investigated by Diseth, Pallesen, Brunborg, and Larsen (2009). Therein, the results of their structural equation modelling (SEM) show that the surface approach to learning has a significant negative effect on the students' performance in the course. However, the deep approach to learning has no significant effect on the students' performance (Diseth et al., 2009). In a similar study, Valadas, Almeida, and Araújo (2016) investigated the role played by approaches to learning between previous and current academic performance of 247 first-year undergraduate students (38% of whom enrolled on science and technology programmes). Using the SEM approach, it was found that previous academic performance predicts both deep and surface approaches to learning. The surface approach to learning predicts the current academic performance, and the deep approach to learning fails to predict the current academic performance (Valadas et al., 2016).

There have also been some findings of no significant correlations between neither deep approach nor surface approach to learning, and students' end of the semester examination scores (Gijbels, Van de Watering, Dochy, & Van den Bossche, 2005; Öhrstedt & Lindfors, 2018). In contrast, Herrmann, Bager-Elsborg, and McCune (2016) reported significant correlations between both deep and surface approaches with grade point average, using 4377 university students. One may argue that the lack of coherence in these findings can be ascribed to the different student populations, diverse fields of study, and lack of uniformity of the measures of performance involved in the studies. However, an exposition of literature with a focus on mathematics learning, that is presented in the next section, refutes this argument. These inconclusive or rather incoherence findings have motivated the present study. Thus, the purpose of the present study is to expose the structural relationship between prior mathematics knowledge, approaches to learning and current mathematics performance among first-year engineering students. We have focused on first-year engineering students learning introductory calculus course in the present study for many reasons. One of these reasons, others are mentioned in the methods section, is that they perform poorly in the course every year. Recent evidence for this claim is clear in a descriptive analysis of grade distribution for autumn semester 2019 that shows 43% failed the course at a Norwegian university. Thus, we aim at addressing the following research question: What are the individual and combined effects of prior mathematics knowledge and approaches to learning on students' performance in an introductory mathematics course?

The remaining part of this paper is arranged as follows: First, a conceptual framework that guides the current study is presented; this includes a theoretical perspective on the approaches to learning, a more focused overview of research that focuses on mathematics learning, and formulations of hypotheses. Second, a section on methodological issues is presented, which is followed by the result section. Third, the results are discussed in a separate section. The paper closes with some remarks, including strengths and potential weaknesses.

CONCEPTUAL FRAMEWORK

Theoretical perspectives on approaches to learning

The concept of approaches to learning was introduced as university students' adopted *processes* before a learning activity that directly affect learning outcomes in the early work of Marton and Säljö (1976). The concept grew rapidly and eventually evolved into approaches to learning (SAL) theory. The SAL theory, as pursued by Marton and Booth (1997), makes an ontological shift from the cognitivist's and constructivist's dualistic view of an individual and the individual's social, cultural, physical and temporal context, when explaining learning. In SAL theory, learning is an activity that leads to an individual's gains in knowledge about the world through experience. The world that is not separated from the individual, neither individually constructed nor culturally imposed. Instead, the world is constituted in a feedback relationship between the individual and the environment as experienced by the individual (Marton & Booth, 1997). As such, in explaining students' learning, SAL theory emphasises "taking the experiences of people seriously and exploring the physical, the social, and the cultural world they experience" (Marton & Booth, 1997, p. 13). The epistemological position in the SAL tradition follows a phenomenographical approach with '*a way of experiencing a meaning that is dialectically intertwined with a structure*' as its unit of analysis (Marton & Booth, 1997). Thus, it is challenging to discern approaches to learning from a few dependent factors such as the previous learning experience, intention to learning, the conception of learning, motives, methods of assessment, and the learning context (Biggs et al., 2001; Iannone, Czichowsky, & Ruf, 2020; Marton & Säljö, 2005).

Given that students experience the world in qualitatively different ways, it is logical to acknowledge the plausibility of different approaches to learning among them. However, a series of empirical studies in the SAL tradition have shown that these diverse approaches to learning can be substantially described using deep and surface approaches (Entwistle, 2005; Marton & Säljö, 2005; Svensson, 2005). The deep approach to learning has been conceptualised as "activities that are appropriate to handling the task so that an appropriate outcome is achieved" (Biggs, 2012, p. 42). Within the context of mathematics learning, appropriate activities entail paying keen attention to the content of the learning material with the intention of proper grasping of the content. In this case, the appropriate learning outcome/performance of a student that adopts a deep approach to mathematics naturally comes as a by-product of his/her understanding of the content and not the primary aim. In contrast, the surface approach to learning has been conceptualised as "activities of an inappropriately low cognitive level, which yields fragmented outcomes that do not convey the meaning of the encounter" (Biggs, 2012, p. 42). Thus, students who adopt surface approaches to learning mathematics are usually driven by the intention to pass the course with minimal effort and therefore concentrate on critical points of the learning materials.

Consistent with the ontological and epistemological claims of the SAL theory, approaches to learning mathematics have been argued to be context-specific rather than being fixed personal characteristics, motivated by intentions, good predictors of students' performance on mathematics tasks (Zakariya, Nilsen, et al., 2020). The variability in the adoption of either deep or surface approaches to learning mathematics has been empirically shown to depend on the nature of mathematics tasks that the students are exposed to (Maciejewski & Merchant, 2016). A mathematics course, such as an introductory first-year calculus, in which the curriculum emphasis is mostly on mastery of techniques of differentiation and integration of functions, has the tendency to encourage adoption of surface approaches to learning. On the other hand, higher-level university mathematics courses such as real analysis, abstract algebra, and numerical analysis of differential equations, with emphases mostly on proofs and conceptual mathematical arguments, have the tendency to encourage adoption of deep approaches to learning. These qualitative differences in students' approaches to learning mathematics are theorised to be related to qualitative differences in the students' learning outcomes on the courses. Thus, it can be argued that approaches to learning mathematics influence the performance of engineering students on mathematics tasks. As such, the following hypothesis is formulated:

> *Hypothesis 1*: There are effects of engineering students' approaches to learning on their performance in a first-year calculus course

Deep and surface approaches to learning and students' performance in mathematics

The situation of the relationships between deep and surface approaches to learning and students' performance in mathematics has not been that different from the previously reported studies in other fields. There is a body of research that supports the plausibility of Hypothesis 1. Even though the study by Mundia and Metussin (2019) shows no significant effect of neither deep approach nor surface approach to learning on achievement of students in mathematics. The debate among researchers on mathematics learning surrounds the question of which of the two approaches (deep or surface) influence students' current performance or is influenced by students' past performance in mathematics? In a study that involves 899 undergraduate students following first-year mathematics classes by Cano et al. (2018), it was found that a deep approach to learning has a positive effect on students' mathematics achievement in the course. In contrast, the effect of the surface approach to learning on the students' mathematics achievement was not significant (Cano et al., 2018). Maciejewski and Merchant (2016) reported mixed findings on the effect of approaches to learning on students' current performance depending on years of study. For students in year-one, it was found that a deep approach to learning predicts students' grades on a mathematics course while the surface approach to learning does not predict the course grades. However, for students in other years of study, the surface approach to learning predicts students' course grades negatively while the deep approach to learning does not predict the course grades (Maciejewski & Merchant, 2016).

One may argue, as did, and subsequently shown empirically by, Maciejewski and Merchant (2016), that the variability in the mathematics tasks exposure across different years of university study could account for the inconsistent prediction of students' course grades by the approaches to learning. However, findings from other mathematics education literature weaken this argument. For instance, Nguyen (2016) investigated the contribution of admission points, approaches to learning, and demographic factors of 616 students' performance in an introductory first-year calculus course, and reported a contrary finding to the one by Maciejewski and Merchant (2016) for year-one students. Therein, it was found that the surface approach to learning has a negative effect on students' performance as measured by course grades in the course. In contrast, a deep approach to learning fails to predict students' performance in a first-year calculus course. Furthermore, there was a significant effect of past performance, as measured by the admission points, on the current performance in the calculus course (Nguyen, 2016). Another related study, though it was conducted using upper primary school students, is a multiple linear regression analysis study by García et al. (2016). They found that the surface approach to learning predicts students' mathematics performance scores with a negative regression weight. On the other hand, the deep approach to learning does not predict students' performance in mathematics examination (García et al., 2016).

The foregoing discussion has exposed some gaps in research that focuses on mathematics learning about the specific effects of either deep or surface approaches to learning mathematics on students' performance in the subject. As such, more studies aimed at filling these gaps are required. The present study using first-year engineering students enrolled in a mathematics course is hoped to contribute to the ongoing debate in research that focuses on mathematics learning. Thus, by drawing mainly on the literature in the current section coupled some tenets of SAL theory that are presented in the theoretical perspectives section, the following additional hypotheses are formulated and will be subjected to testing in the present study.

> *Hypothesis 2*: There are effects of prior mathematics knowledge on engineering students' deep and surface approaches to learning mathematics.
>
> *Hypothesis 3*: Both the deep and surface approaches to learning mediate the effect of engineering students' prior mathematics knowledge on their current performance in a first-year calculus course.

It is important to remark that the formulation of hypotheses 2 and 3 relies mainly on previous studies as it is typical of practices when using a conceptual framework. As rightly puts by Eisenhart (1991), a conceptual framework is "a skeletal structure of justification, rather than a skeletal structure of explanation based on logic (i.e., formal theory) or accumulated experience (i.e. practitioner knowledge)", (p. 209). As such, a researcher can combine a formal theory with previous studies to formulate the research hypotheses which is contrary to an adoption of a theoretical framework in which case the researcher is confined to the hypotheses of the formal theory (Lester, 2010). Figure 1 presents a hypothesised relationship between the variables.

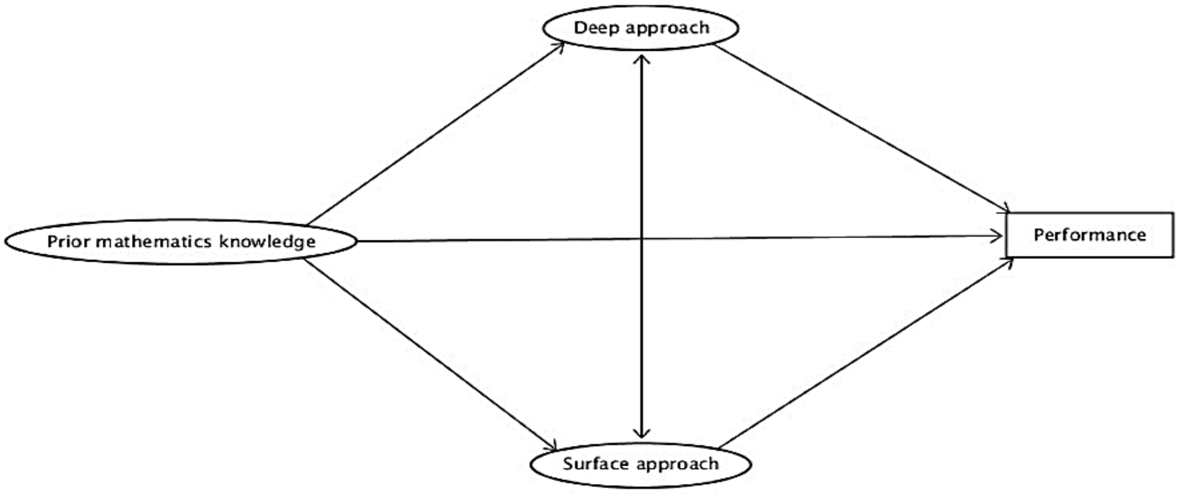


Figure 1. A hypothesised model of relationship between the research constructs

METHODS

Sample of the study

The present study adopts a cross-sectional research design with a focus on first-year engineering students enrolled on an introductory calculus course at a university in Norway during the autumn 2019 semester. The sample is convenient because of the strict rules on individual consent for studies in Norway. First-year university students are chosen for the present study to avoid conflicting findings, as reported by Maciejewski and Merchant (2016) when students from other years of study are included in their study. Also, it is assumed that pre-university mathematics knowledge may be assessed adequately in the first year of study. Further, these students encounter severe challenges in passing the first-year introductory calculus course. Perhaps, it is due to their newness to the university system. Consistent with the task-specificity of approaches to learning, as postulated by the SAL theory, we focus on students following a course. As such, we restrict our sample to engineering students because they form the largest population of students following the same mathematics course at the university. A valid sample of 188 students (25% females) with a mean age of 22 years participated in the present study, which amounts to 63% of the first-year students' enrolment in the course. Moreover, these engineering students have been previously characterised to study mathematics with predominantly deep and surface approaches (Authors, 2019a).

Measures

Two research instruments are used in the present study and the final cumulative examination scores of students are taken as a measure of current mathematics performance on the introductory calculus course. The first research instrument is a measure of prior mathematics knowledge, and the second research instrument is a measure of approaches to learning mathematics. Full descriptions of these instruments are presented in the forthcoming sections.

Prior mathematics knowledge

We adopt a Norwegian national mathematics test (NNMT) that was administered at the time of the study (i.e., autumn 2019) as a proxy to measure prior mathematics content knowledge of the students. The NNMT examines basic knowledge of mathematics with a focus on the upper secondary curriculum. The Norwegian Mathematical Council administers this test to first-year university students across Norway with an intention to monitor the level of preparedness of the students' in mathematics for university education. It is assumed that the NNMT will be most appropriate as a proxy for assessing students' mathematics knowledge in the present study. This is because the test draws on basic mathematics content knowledge that is common to the upper secondary school mathematics curriculum, and it has been consistently used for the past three decades. In addition to the preliminary section, the NNMT contains 16 stem questions on mathematics tasks, and some questions have two or three parts to make a total of 22 items. The NNMT has been subjected to tests of item quality (e.g., difficulty and discrimination indices), and reliability (Authors, 2020a). It was found to possess unidimensionality, appropriate construct validity, and a reliability coefficient of .90. Following the recommendations of the study by Authors (2020a), scores of students on only seventeen items that meet an acceptable level of both difficulty and discrimination indices, are used in the present study. The seventeen items capture simple manipulations of fractions, ratios and percentages, and some short answer word problems.

Approaches to learning mathematics

We adopted a Norwegian version of the two-factor approaches to learning questionnaire (R-SPQ-2F) to assess students' approaches to learning mathematics (Authors, 2020b). The Norwegian version of the R-SPQ-2F (henceforth, R-SPQ-2F) was adopted in the present study for several reasons. First, it was conceptualised and operationalised to expose approaches to learning based on the postulates of the SAL theory (Biggs et al., 2001). Second, it has been validated using data from students on a mathematics course, which is in line with the task-specificity of the approaches to learning (Authors, 2019a). Third, it has only 19 items (ten items on deep and nine items on surface approaches to learning mathematics) as compared to the study skills inventory for students (52 items), or the revised approaches to studying inventory (24 items). It is available in Norwegian, which takes care of any context, language, or cultural dependence of the construct it is purported to measure. Example of an item on the deep approach (a reliability coefficient of .81) subscale of the R-SPQ-2F is "I test myself on important topics until I understand them completely," while "I do not find my course very interesting so I keep my work to the minimum" is an example of an item on a surface approach subscale (a reliability coefficient of .81) (Authors, 2020b). Students are meant to self-report their dispositions to items on the R-SPQ-2F using a five-point Likert scale from "never or only rarely true of me" to "always or almost always true of me". The R-SPQ-2F has been found to possess high construct validity (Authors, 2020b), and discriminant validity (Authors, 2019b).

Current mathematics performance of students

The current mathematics performance of students (henceforth, performance) was measured using the final scores of each participating student in the introductory calculus course. The final examination is an individual written in-class examination that lasts for four hours. It is assumed that these scores are reflective of students' performance in the course.

Data curation, ethics, and analysis

Data curation and ethical considerations

The NNMT and the R-SPQ-2F are prepared using online SurveyXact and distributed via email to collect data used for the present study at two instances. Instance one of the data collection took place at the beginning of the autumn 2019 semester in which the students sat for the online NNMT in classrooms under the supervision of their class teachers. It took up to 40 minutes to complete the test. It is assumed that the early weeks of the semester is the best time to administer the NNMT for a valid assessment of the prior mathematics knowledge. The second data collection point occurred towards the last week of the semester using the R-SPQ-2F to assess students' ongoing approaches to learning the subject. Meanwhile, before the first data collection point, the students were informed of the purpose of study during a class visit by the researchers, and their consent was sought. They gave their voluntary consent (written and via digital channel) to take part in the study and to authorize the researchers to access their final grades in the course via the university students' examination office.

Data analysis

The collected data were screened and found to contain no out-of-range value on any of the research variables. Some data were missing at random, especially on the NNMT and exam scores as a result of some discrepancies in the number of students who sat for the two tests. These missing data are minimal (between 12-20 cases) and therefore assumed to have no substantial effect on the results of the analysis. The data were analysed using an item-level SEM approach. This involves a two-step method of firstly evaluating the measurement models for the NNMT and the R-SPQ-2F before proceeding to evaluating the structural model in line with the best practice in SEM literature (Byrne, 2012). The SEM approach was used because of its satisfactory performance in predictions over linear regression models and path analyses (Bollen & Pearl, 2013). Data from both the NNMT and the R-SPQ-2F are categorical because of the correct/incorrect and the Likert scale response formats used in the scoring of the instruments, respectively. As such, the weighted least square mean and variance adjusted (WLSMV) estimator with theta parameterization was used in the evaluation of the models. We assess the local fits of the models by looking at the significant level of the item factor loadings, item residuals, and standard errors. The global fits of the models were assessed with a combination of fit indices, as recommended in the literature for an acceptable fit. First, the ratio of chi-square value to the degree of freedom with a magnitude of less than three was used. This ratio was combined with a value of close to or greater than .90 for both the TLI-Tucker-Lewis index and the CFI-comparative fit index (Bentler, 1990). Further, an RMSEA-root mean square error of approximation, the value of less than .06, was used to judge a good fit (Hu & Bentler, 1999). It is important to remark that we do not use chi-square value to assess the global fit of the models because of its sensitivity to large sample size (Brown, 2015). Instead, we use its ratio to the degree of freedom. As such, we do not include its p-values in our results.

RESULTS

Step one: Evaluation of measurement models

A one-factor measurement model was evaluated for the NNMT by fitting the seventeen items on the test while constraining the variance of the latent factor (prior mathematics knowledge) to 1. The constrained latent factor variance to 1 ensures model identification and allows the factor loadings of each of the items to be freely estimated. Further, a two-factor measurement model was evaluated for the R-SPQ-2F. Following the recommendations by Authors (2020b), item10 and item19 that are hypothesised to expose surface approach to learning were cross-loaded on the deep approach subscale, and an error covariance was allowed between item15 and item18 on the surface approach subscale. The models for the NNMT and the R-SPQ-2F were evaluated separately. Some selected global fit statistics of these models are presented in Table 1.

Table 1.

Global fit statistics of the NNMT and R-SPQ-2F measurement models

Global fit statistics	NNMT model	R-SPQ-2F model
Chi-square		
Estimate (χ^2)	143.793	283.336
Degrees of freedom (df)	119	148
χ^2 / df	1.21	1.91
RMSEA		
Estimate	.043	.070
90 percent confidence interval	[< .001, .066]	[.058, .082]
Probability RMSEA <= .05	.676	.005
TLI/CFI		
TLI	.936	.895
CFI	.944	.909

The results presented in Table 1 show that both the NNMT and the R-SPQ-2F have acceptable measurement model fits as indicated by global fit statistics. The ratios of the chi-square values to the degrees of freedom are less than the recommended value criterion for an acceptable fit. The RMSEA value of the NNMT model with a non-significant 90 percent confidence interval shows that there is a substantial agreement between the NNMT model and the collected data. Even though the RMSEA value of the R-SPQ-2F model is higher than the .06 recommended value by Hu and Bentler (1999), its 90 percent confidence interval includes .06 and, as such, considered for an acceptable fit (MacCallum, Browne, & Sugawara, 1996). The TLI and the CFI values of both NNMT and the R-SPQ-2F models are within the ranges suggested for an acceptable fit. In addition to the global fit statistics presented in Table 1, the local fits statistics show acceptable values. The factor loadings are significant and moderately high for each item in the two models, the item residuals and the standard errors are within acceptable ranges. Both the global and local fits statistics of the NNMT and R-SPQ-2F models are suggestive of the plausibility of the models. Thus, we proceed to the analysis of the structural evaluation of the models.

Step two: Evaluations of structural models (hypothesis testing)

In order to test the hypotheses of the present study and by extension to address the research question, we evaluate the hypothesised model of relationship between the research constructs (Figure 1). The results of selected global fit statistics of the structural model are presented in Table 2, and Figure 2 shows the parameter estimates of the model.

Table 2.

Selected global fit statistics of the relationship between prior mathematics knowledge, approaches to learning and performance

Global fit statistics	Structural model
Chi-square	

Estimate (χ^2)	717.548
Degrees of freedom (df)	621
χ^2 / df	1.16
RMSEA	
Estimate	.029
90 percent confidence interval	[.017, .038]
Probability RMSEA <= .05	1.000
TLI/CFI	
TLI	.929
CFI	.934

The global fit statistics presented in Table 2 show an acceptable global model fit of the relationship between the research constructs. This acceptable fit is deduced from the fact that all the global fits statistics are within the recommended values. The results presented in Table 2 guarantee the plausibility of the standardized model estimates of effects (standard errors – S.E.s) that are presented in Figure 2.

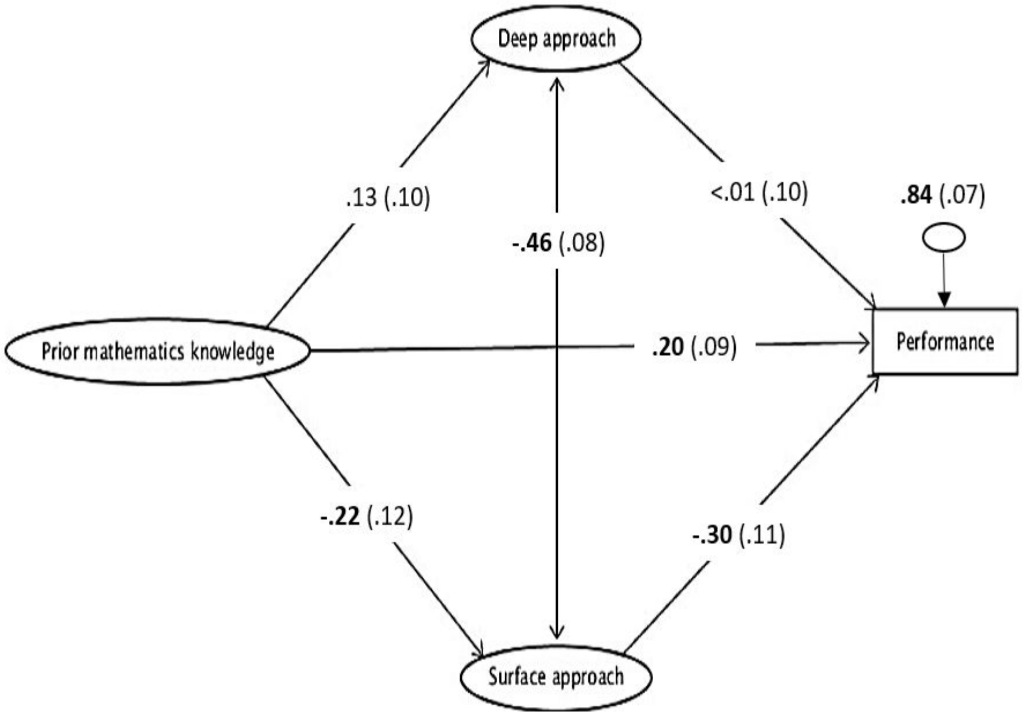


Figure 2. The evaluated structural model of the relationship between prior mathematics knowledge, approaches to learning, and students' performance. The significant effect paths at p<.05 are in boldface. A full figure that contains all the items and the associated parameter estimates is available in the appendix.

The results presented in Figure 2 show that the surface approach to learning has a standardised significant negative effect (β = -.30, S.E. = .11, $p < .05$) on the performance of students. In contrast, the deep approach to learning has negligible effect ($\beta < .01$, S.E. = .10, $p > .05$) on the performance. This finding substantiates the plausibility of Hypothesis 1 of the present study. As such, approaches to learning influence students' performance but only through the surface approach. The double-headed arrow between the deep and the surface approaches shows a standardised significant correlation coefficient ($r = -.46$, S.E. = .08, $p < .05$) between errors involved in measuring the two latent constructs. This error covariance is acknowledged and included in the model to enhance reliable model estimates. It is also shown in Figure 2 that the standardised effect of the prior mathematics knowledge on the surface approach to learning is significant and negative ($\beta = -.22$, S.E. = .12, $p < .05$). In contrast, there is no significant standardised effect of the prior mathematics knowledge on the deep approach to learning ($\beta = .13$, S.E. = .10, $p > .05$). These findings partly confirm the plausibility of Hypothesis 2 by establishing the effects of the prior mathematics knowledge on approaches to learning. Further, the findings show that the surface approach to learning is the only predictable approach by the prior mathematics knowledge of students. Moreover, the results presented in Figure 2 show that prior mathematics knowledge has a significant effect ($\beta = .20$, S.E. = .09, $p < .05$) on the students' performance. The small arrow with a significant estimate of .84 (S.E. = .07, $p < .05$) is the residual variance involved in the prediction of performance by both the prior mathematics knowledge and the approaches to learning.

Mediation analysis results

We conducted a partial mediation analysis on the structural model presented in Figure 2 in order to investigate the total effect and specific mediating effects of the prior mathematics knowledge through the approaches to learning on performance.

Table 3. Mediation analysis results from prior mathematics knowledge through approaches to learning to performance

Effects from prior mathematics knowledge to performance	Estimate	Lower 5%	Upper 5%
Total effect	.268	.109	.427
Specific indirect effect (Surface approach)	.136	.068	.147
Specific indirect effect (Deep approach)	< .001	-.021	.020

The presented results in Table 3 of the partial mediation analysis show that the prior mathematics knowledge has a total effect of $\beta = .268$ [.109, .427] on performance. The 95% percent confidence interval of this estimate using bootstrapping [.109, .427] shows that the total effect can go as high as .427. Since the confidence does not include 0, we conclude that the total effect size is significant. The specific standardised mediating effect of the prior mathematics knowledge through the surface approach to learning on performance is $\beta =.136$ [.068, .149]. Since the 95% confidence interval of this estimate does not include 0, we conclude that this mediating effect is significant. In contrast, the specific standardised mediating effect of the prior mathematics knowledge through the deep approach to learning on performance is $\beta < .001$[-.021, .020]. The 95% confidence interval of this estimate includes 0. Thus, the standardised mediating effect is not significant. Therefore, it can be deduced from the mediation analysis results that, of the two approaches to learning, only the surface approach to learning partially mediates the effect of the prior mathematics knowledge on performance, which partially contradicts Hypothesis 3.

DISCUSSION

The concept of approaches to learning has been thoroughly articulated and operationalised by the SAL theory. The approaches adopted by the students to learn explain some qualitative differences in their learning outcomes (Marton & Säljö, 2005). However, there have been heated debates among education researchers on the contributions of either deep or surface approaches to the learning outcomes on mathematics tasks. Perhaps, these debates were generated by the inconclusive findings of the empirical studies reported with this intention. As such, attempts are made in the present study to disentangle the complicated relationship between approaches to learning mathematics, previous and current students' performance on mathematics tasks. The first step of our analysis involves evaluations of the measurement models of the NNMT and the R-SPQ-2F using data collected from first-year engineering students enrolled on an introductory calculus course. The results of this analysis confirm the factor structures of these instruments, as postulated in the literature (e.g., Authors, 2020b). The second step is a structural evaluation of a hypothesised model of the relationship between these constructs and performance. The ensuing findings are discussed in the next section.

The effect of approaches to learning on performance

The findings of the present study show that there is a significant negative effect of the surface approach to learning on performance. In contrast, the deep approach to learning has no significant effect on performance. One may interpret these findings to mean that students who adopt surface approaches to learning mathematics have low scores (grades) in the introductory calculus course. However, there is no evidence in the present study to justify a substantial relationship between students who adopt deep approaches to learning mathematics and their performance in the course. As such, one may discourage adoption of the surface approach to learning, but that does not mean encouragement of the deep approach to learning either. It is less surprising that the deep approach to learning does not influence students' performance in the present study, given that similar findings have been reported, elsewhere (e.g., García et al., 2016). Even though some researchers have reported a substantial contribution of the deep approaches to learning to students' performance in mathematics (e.g., Cano et al., 2018). Therefore, based on the SAL theory, a major contribution of the present study is an identification of the surface approach to learning as the only predictor of performance among first-year engineering students. This contribution fits very well among the existing research in the field (Mundia & Metussin, 2019; Nguyen, 2016).

The effects of prior mathematics knowledge on approaches to learning and performance.

It was also found in the present study that prior mathematics knowledge has significant positive and negative effects on the students' performance and the surface approach to learning, respectively. These findings can be interpreted to mean, students who have high scores on the NNMT performed very well in the introductory calculus course while those students who have low scores on the NNMT are associated with the surface approaches to learning. The fact that prior mathematics knowledge predicts the surface approach to learning negatively is consistent with the postulates of the SAL theory and corroborates the findings of previous studies (e.g., Valadas et al., 2016). The positive relationship between the prior mathematics knowledge and the students' performance is expected, given the level of precision offered by the SEM approach used in scoring the former. This finding corroborates the findings of previous studies that have reported a similar relationship between the prior and current performance of students (Diseth et al., 2009; Valadas et al., 2016). Further, it was found that the positive effect of prior mathematics knowledge on the deep approach to learning is not significant. This finding shows that students who scored high on the NNMT also scored high on the deep approach subscale of the R-SPQ-2F. However, the strength of the relationship between the scores of students on the two variables is not enough to substantiate any sizeable relationship between them. Even though the finding appears not to agree with the postulates of the SAL theory, it does support the finding of no significant relationship between prior academic achievement and the deep approach to learning reported by Diseth et al. (2009).

Mediating effects of the approaches to learning

An essential finding of the present study comes from the results of the partial mediation analysis. It is a partial mediation because the direct effect of prior mathematics knowledge on performance was studied simultaneously with its indirect effect through approaches to learning, unlike a full mediation analysis where the direct effect from the predictor variable to the outcome variable will not be included in the analysis (Rucker, Preacher, Tormala, & Petty, 2011). The finding shows that the effect of prior mathematics knowledge on performance is mediated by the surface approach to learning. In contrast, the deep approach to learning fails to mediate this effect. An interpretation of this finding is that, given that the direct effect is controlled for, students with high scores on the NNMT and low scores on the surface approach subscale performed better in the introductory calculus course than those students who have low scores on the NNMT and high on the surface approach subscale. However, we found no substantial evidence to make a similar interpretation for the deep approach subscale. We claim that our finding on the mediating effect of approaches to learning as intervening factors between previous and current performance in mathematics is a new observation in research with a focus on mathematics learning. An extensive search of the literature corroborates this claim. The authors could only find few studies in other fields, for instance, the study by Valadas et al. (2016), who found no substantial mediating effect of neither deep nor surface approaches to learning between prior knowledge and academic success among first-year university students. One may argue that the reason why Valadas et al. (2016) could not detect the mediating effect is that they measured academic success by a self-reported number of failed courses by the students. In the present study, given that we used examination scores that are common to every student for measuring performance, it is expected that our finding on the mediating effect is closer to reality than theirs.

CONCLUDING REMARKS

We conclude this article by highlighting some perceived strengths of the study, potential limitations, and recommendations for future studies. First, the research instruments used in the present study are carefully prepared based on theoretical and empirical evidence. It has been argued in the literature (e.g., Zakariya, 2020) that the validity of findings from quantitative research depends mainly on the quality of research instruments. As such, the specially tailored measures of the prior mathematics knowledge and the approaches to learning mathematics constitute a strength of the present study. Second, unlike the correlation/regression studies that are usually reported in the quantitative research with a focus on mathematics learning (e.g., Nguyen, 2016), the use of SEM offers some strength to the validity of our findings. The ability of the SEM approach to model error covariance, item cross-loading, and to study the relationship between latent factors demonstrates SEM's superiority over path analysis and multiple linear regression (Bollen & Pearl, 2013). By implications, the findings of the present study underscore the importance of discouraging engineering students from capitalising on the surface approach to learning mathematics. As such, university teachers, curriculum planners, and other stakeholders in the teaching of mathematics to engineering students are charged with the responsibility of guiding the students towards a good approach to learning the subject.

Despite the strengths of the present study, some limitations are recognised. The first limitation may stem from the non-inclusion in our model, factors such as self-efficacy, the conception of and attitudes towards mathematics, gender, and age that could affect students' performance. The non-inclusion of these factors in our model could, perhaps, explain the sizeable residual variance of .84 that was found. That means only 16% of the students' performance variability is explained by prior mathematics knowledge and approaches to learning. We recommend further investigation with this intention. It is acknowledged that the self-report questionnaire (R-SPQ-2F) used in the present study carries with itself some weaknesses. It does not cover all aspects of the students' approaches to learning, and some students might fake their dispositions or be too shy to reveal their actual thoughts while answering the questionnaire. The present study does not account for all these specific shortcomings of the R-SPQ-2F, which might constitute a threat to the findings. However, with the inclusion of measurement errors, we hope that some of these shortcomings are taken care of in the present study. Though, restricting our sample to engineering students in their first year of study increases the precision of the model estimates but may limit generalisation beyond this student's population. Further, the sample size appears to be small. As such, we recommend future replications of the present study in a large and diverse student population.

DECLARATIONS

Funding

No funding information is available.

Conflicts of interest/Competing interests

The authors declare no conflict of interest

Availability of data and material

The anonymous data used for the present study are available upon request from the corresponding author

REFERENCES

Alkhateeb, H. M., & Hammoudi, L. (2006). Attitudes toward and approaches to learning first-year university mathematics. *Perceptual and motor Skills, 103*, 115-120. doi:10.2466/pms.103.1.115-120

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Biggs, J. B. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Educational Research Journal, 63*, 3-19. doi:10.1111/j.2044-8279.1993.tb01038.x

Biggs, J. B. (2012). What the student does: Teaching for enhanced learning. *Higher Education Research & Development, 31*(1), 39-55. doi:10.1080/07294360.2012.642839

Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149. doi:10.1348/000709901158433

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation. In S. L. Morgan (Ed.), *Handbooks of Sociology and Social Research* (pp. 301-328). Dordrecht: Springer.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, London: The Guilford Press.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge, Taylor & Francis Group.

Cano, F., Martin, A. J., Ginns, P., & Berbén, A. B. G. (2018). Students' self-worth protection and approaches to learning in higher education: Predictors and consequences. *Higher Education, 76*(1), 163-181. doi:10.1007/s10734-017-0215-0

Diseth, Å. (2007). Students' evaluation of teaching, approaches to learning, and academic Achievement. *Scandinavian Journal of Educational Research, 51*(2), 185-204. doi:10.1080/00313830701191654

Diseth, Å., Pallesen, S., Brunborg, G. S., & Larsen, S. (2009). Academic achievement among first semester undergraduate psychology students: the role of course experience, effort, motives and learning strategies. *Higher Education, 59*(3), 335-352. doi:10.1007/s10734-009-9251-8

Eisenhart, M. A. (1991). *Conceptual frameworks for research circa 1991: Ideas from a cultural anthropologist; implications for mathematics education researchers.* Paper presented at the 13th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Blacksburg, Virginia, USA.

Entwistle, N. (1997). Contrasting Perspectives on Learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning : Implications for teaching and studying in higher education*. Edinburgh: Scottish Academic Press.

Entwistle, N. (2005). Contrasting perspectives on learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 106-125). Edinburgh: The University of Edinburgh, Centre for Teaching, Learning and Assessment.

García, T., Rodríguez, C., Betts, L., Areces, D., & González-Castro, P. (2016). How affective-motivational variables and approaches to learning predict mathematics achievement in upper elementary levels. *Learning and Individual Differences, 49*, 25-31. doi:10.1016/j.lindif.2016.05.021

Gijbels, D., Van de Watering, G., Dochy, F., & Van den Bossche, P. (2005). The relationship between students' approaches to learning and the assessment of learning outcomes. *European Journal of Psychology of Education, 20*(4), 327-341. doi:10.1007/BF03173560

Guo, J., Yang, L., & Shi, Q. (2017). Effects of perceptions of the learning environment and approaches to learning on Chinese undergraduates' learning. *Studies in Educational Evaluation, 55*, 125-134. doi:10.1016/j.stueduc.2017.09.002

Herrmann, K. J., Bager-Elsborg, A., & McCune, V. (2016). Investigating the relationships between approaches to learning, learner identities and academic achievement in higher education. *Higher Education, 74*(3), 385-400. doi:10.1007/s10734-016-9999-6

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Iannone, P., Czichowsky, C., & Ruf, J. (2020). The impact of high stakes oral performance assessment on students' approaches to learning: a case study. *Educational Studies in Mathematics*. doi:10.1007/s10649-020-09937-4

Lester, F. K., Jr. (2010). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. In B. Sriraman & L. English (Eds.), *Theories of Mathematics Education, Advances in Mathematics Education* (pp. 67-85). Berlin Heidelberg: Springer-Verlag.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149. doi:10.1037/1082-989X.1.2.130

Maciejewski, W., & Merchant, S. (2016). Mathematical tasks, study approaches, and course grades in undergraduate mathematics: a year-by-year analysis. *International Journal of Mathematical Education in Science and Technology, 47*(3), 373-387. doi:10.1080/0020739x.2015.1072881

Marton, F., & Booth, S. A. (1997). *Learning and awareness*. Hillsdale, NJ: Lawrence Erlbaum.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning I: Outcome and process. *British Journal of Educational Psychology, 46*, 4-11. doi:10.1111/j.2044-8279.1976.tb02980.x

Marton, F., & Säljö, R. (2005). Approaches to Learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning : Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 39-58). Edinburgh: The University of Edinburgh, Centre for Teaching, Learning and Assessment.

Mji, A. (2000). Conceptions of and approaches to learning mathematics. *Psychological Reports, 87*, 971-972. doi:10.2466/pr0.2000.87.3.971

Mundia, L., & Metussin, H. (2019). Exploring factors that improve mathematics achievement in Brunei. *Studies in Educational Evaluation, 60*, 214-222. doi:10.1016/j.stueduc.2018.10.003

Nguyen, T. M. (2016). Learning approaches, demographic factors to predict academic outcomes. *International Journal of Educational Management, 30*(5), 653-667. doi:10.1108/ijem-06-2014-0085

Öhrstedt, M., & Lindfors, P. (2018). Linkages between approaches to learning, perceived stress and expected and actual academic outcomes among first-semester psychology students. *Journal of Further and Higher Education, 42*(1), 116-129. doi:10.1080/0309877x.2016.1206856

Rozgonjuk, D., Kraav, T., Mikkor, K., Orav-Puurand, K., & Täht, K. (2020). Mathematics anxiety among STEM and social sciences students: the roles of mathematics self-efficacy, and deep and surface approach to learning. *International Journal of STEM Education, 7*(1). doi:10.1186/s40594-020-00246-z

Rucker, D., Preacher, K., Tormala, Z., & Petty, R. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*, 359-371. doi:10.1111/j.1751-9004.2011.00355.x

Svensson, L. (2005). Skill in learning and organising knowledge. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (3rd (Internet) ed., pp. 59-71). Edinburgh: The University of Edinburgh, Centre for Teaching, Learning and Assessment.

Tait, H., Entwistle, N. J., & McCune, V. (1998). ASSIST: a re-conceptualisation of the approaches to studying inventory. In C. Rust (Ed.), *Improving students as learners* (pp. 262-271). Oxford: Oxford Brookes University.

Trigwell, K., Ellis, R. A., & Han, F. (2012). Relations between students' approaches to learning, experienced emotions and outcomes of learning. *Studies in Higher Education, 37*(7), 811-824. doi:10.1080/03075079.2010.549220

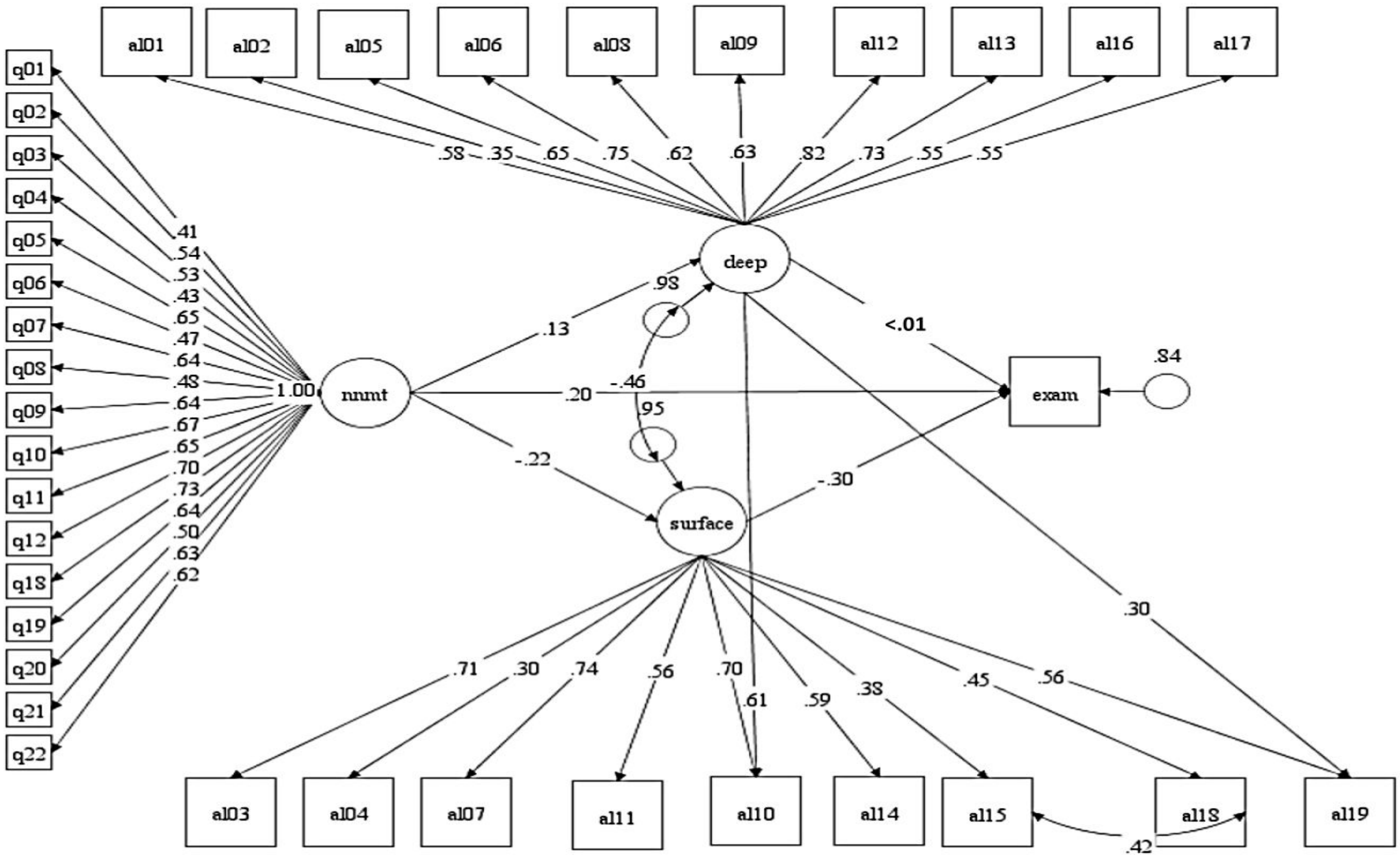
Valadas, S. T., Almeida, L. S., & Araújo, A. M. (2016). The mediating effects of approaches to learning on the academic success of first-year college students. *Scandinavian Journal of Educational Research, 61*(6), 721-734. doi:10.1080/00313831.2016.1188146

Zakariya, Y. F. (2020). Investigating some construct validity threats to TALIS 2018 teacher job satisfaction scale: Implications for social science researchers and practitioners. *Social Sciences, 9*(4). doi:10.3390/socsci9040038

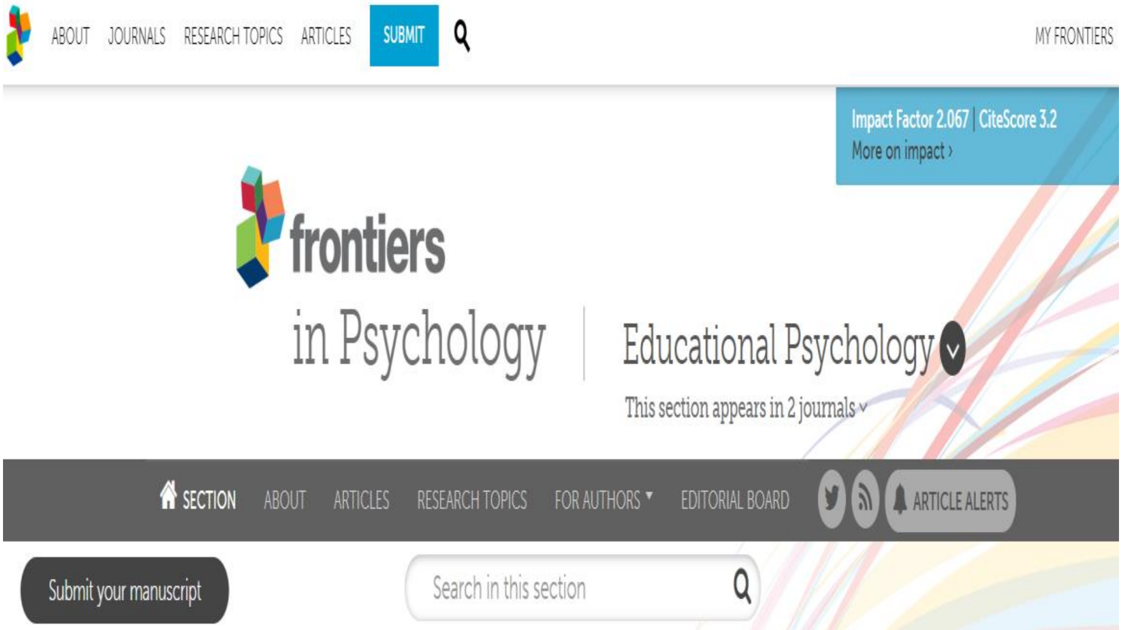
Zakariya, Y. F., Bjørkestøl, K., Nilsen, H. K., Goodchild, S., & Lorås, M. (2020). University students' learning approaches: an adaptation of the revised two-factor study process questionnaire to Norwegian. *Studies in Education Evaluation*, 100816. doi:10.1016/j.stueduc.2019.100816

Zakariya, Y. F., Nilsen, H. K., Goodchild, S., & Bjørkestøl, K. (2020). Self-efficacy and approaches to learning mathematics among engineering students: Empirical evidence for potential causal relations. *International Journal of Mathematical Education in Science and Technology*, 1-15. doi:10.1080/0020739X.2020.1783006

Appendix: The full structural model of effects of prior mathematics knowledge, and approaches to learning on students' performance.



Paper VIII



Zakariya, Y. F. (2021). Self-efficacy between previous and current mathematics performance of undergraduate students: an instrumental variable approach to exposing a causal relationship. *Frontiers in Psychology*. 11:556607. doi:10.3389/fpsyg.2020.556607.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fpsyg.2020.556607

Self-Efficacy Between Previous and Current Mathematics Performance of Undergraduate Students: An Instrumental Variable Approach to Exposing a Causal Relationship

*Yusuf F. Zakariya**

Department of Mathematical Sciences, University of Agder, Kristiansand, Norway

Purpose: Self-efficacy has been argued theoretically and shown empirically to be an essential construct for students' improved learning outcomes. However, there is a dearth of studies on its causal effects on performance in mathematics among university students. Meanwhile, it will be erroneous to assume that results from other fields of studies generalize to mathematics learning due to the task-specificity of the construct. As such, attempts are made in the present study to provide evidence for a causal relationship between self-efficacy and performance with a focus on engineering students following a mathematics course at a Norwegian university.

Method: The adopted research design in the present study is a survey type in which collected data from first-year university students are analyzed using structural equation modeling with weighted least square mean and variance adjusted (WLSMV) estimator. Data were generated using mainly questionnaires, a test of prior mathematics knowledge, and the students' final examination scores in the course. The causal effect of self-efficacy was discerned from disturbance effects on performance by using an innovative instrumental variable approach to structural equation modeling.

Results: The findings confirmed a significant direct effect of the prior mathematics knowledge test ($\beta = 0.52$, SE = 0.01, $p < 0.001$) on self-efficacy, a significant direct effect ($\beta = 0.43$, SE = 0.19, $p = 0.02$) of self-efficacy on performance, and a substantial mediating effect ($\beta = 0.22$, SE = 0.10, $p = 0.03$) of self-efficacy between a prior mathematics knowledge test and performance. Prior mathematics knowledge and self-efficacy explained 30% variance of the performance. These findings are interpreted to be substantial evidence for the causal effect of self-efficacy on students' performance in an introductory mathematics course.

Conclusion: The findings of the present study provide empirically supports for designing self-efficacy interventions as proxies to improve students' performance in university mathematics. Further, the findings of the present study confirm some postulates of Bandura's agentic social cognitive theory.

Keywords: self-efficacy, prior mathematics knowledge, undergraduate learning, causal model analysis, instrumental variable approach

Edited by:
Douglas F. Kauffman,
Medical University of the Americas – Nevis, United States

Reviewed by:
Ove Edvard Hatlevik,
Oslo Metropolitan University, Norway
Michael S. Dempsey,
Boston University, United States

Correspondence:
Yusuf F. Zakariya
yusuf.zakariya@uia.no

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *28 April 2020*
Accepted: *28 December 2020*
Published: *18 January 2021*

Citation:
Zakariya YF (2021) Self-Efficacy Between Previous and Current Mathematics Performance of Undergraduate Students: An Instrumental Variable Approach to Exposing a Causal Relationship. Front. Psychol. 11:556607. doi: 10.3389/fpsyg.2020.556607

INTRODUCTION

There has been a growing interest in research on students' affective factors and their contributions to learning outcomes at all levels of education. Apart from the fact that some of these affective factors, e.g., self-efficacy, satisfactorily predict students' performance, an explanation for the growing interest may be ascribed to the ease of developing interventions that influence such factors (Czocher et al., 2019). For instance, perceived self-efficacy, which has been conceptualized as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3), was shown to predict academic achievement better than intelligence test scores, measures of self-esteem, and personal traits among school children (Zuffianò et al., 2013; Özcan and Eren Gümüş, 2019). With regards to the learning outcomes in undergraduate mathematics, perceived self-efficacy was found to be a better predictor of performance than the usefulness of mathematics, prior mathematics knowledge, self-concept (Pajares and Miller, 1994), mathematics anxiety, and mental ability (Pajares and Kranzler, 1995). A high sense of self-efficacy has also been linked with the adoption of deep approaches to learning, high learning motivation, positive attitude toward mathematics. In contrast, a low sense of self-efficacy has been linked with the adoption of surface approaches to learning, high mathematics anxiety, and low interest in mathematics (Bandura, 1997; Rozgonjuk et al., 2020; Zakariya et al., 2020b). More recently, Schukajlow et al. (2019) demonstrate an approach through which constructing multiple solutions to real-life problems can be used as an intervention to influence students' self-efficacy in mathematics. Student-centered instructional methods have also been linked with high self-efficacy (Lahdenperä et al., 2019).

Even though the relationship between self-efficacy and students' performance has been widely studied, little is known about the causal effect of the former on the latter as it concerns the learning of university mathematics. The available studies on self-efficacy with a focus on university mathematics are either relatively old (e.g., Hackett and Betz, 1989; Pajares and Miller, 1994), utilized regression models which make it difficult to evaluate causal hypotheses between self-efficacy and students' performance in mathematics (e.g., Peters, 2013), or do not account for confounding factors in their structural models (e.g., Roick and Ringeisen, 2018). By a causal effect, the author means, if A is a cause of B then at least all the following conditions are satisfied: (1) A temporarily precedes B, i.e., data on A are collected before data on B or A is theorized to happen before B; (2) There is a substantial correlation between A and B; (3) There should not be a third variable C that explains the relationship between A and B (Antonakis et al., 2010). The third condition is the most difficult to meet, especially in non-experimental research. Such variable C will always exist. The most important question is how well a researcher can control it? Among the several attempts that have been shown empirically to yield satisfactory performance in controlling for an extraneous variable, such as C in non-experimental research, is the use of instrumental variable approach (Antonakis et al., 2010; Bollen, 2019). The basic idea of the instrumental variable approach is to find a fourth variable called an *instrument* that satisfies some properties (which will be explained in the "Materials and Methods" section) and use it to discern the actual effect of A on B from any confounding effects of C (Greenland, 2000; Bollen, 2019).

As such, the primary purpose of the present study is to investigate the causal effects of perceived self-efficacy on the current students' performance in mathematics among engineering students with an application of the innovative instrumental variable approach to modeling. Further, the effects of prior mathematics knowledge on the perceived self-efficacy and the current students' performance are also investigated. An advantage of using the innovative instrumental variable approach in exposing these causal effects lies in a fact that reliable estimates of effects can be justified. Despite the wide application of the instrumental variable approach among epidemiologists and econometricians (Antonakis et al., 2010), it is innovative in the present study because the author is not aware of its previous use in mathematics education research. It is the opinion of the author that policymakers, researchers, and education stakeholders are more interested in studies that explore answers to questions on what brings about improved students' performance and to what extent? Rather than, in studies that focus on correlations between variables whose findings are either complicated to interpret or beset by unclear conclusions (Pajares and Miller, 1994). The present study, therefore, attempts to address the following research question: What are the direct and indirect causal effects of prior mathematics knowledge and perceived self-efficacy on performance in mathematics among engineering students? The author draws on both theoretical and analytical perspectives to address this question. The statistical analyses in the present article are moderately advanced and up to date. However, the author has deliberately chosen a simple language of presentation with less mathematical abstractions to make the findings more accessible.

The remaining part of the present article is organized as follows: An overview of a theoretical perspective which leads to the formulation of research hypotheses is presented in the next section. Next is the "Materials and Methods" section where research methodological related issues are presented. The fourth section presents analyses and results. The major findings are discussed in the fifth section, including potential limitations and recommendations for further studies. Finally, the article closes with some remarks.

CONCEPTUAL FRAMEWORK

Perceived self-efficacy is firmly rooted in the agentic social cognitive theory (henceforth, social cognitive theory) as propagated by Albert Bandura in his decades of work on the theory (Bandura, 2001, 2012). Bandura, dissatisfied with some ontological and epistemological claims of traditional cognitive theory (cognitive theory), developed the social cognitive theory. The ontological paradigm shift from the cognitive theory lies in a rejection of dualism between personal agent and object of actions. Reciprocal determinism is an epistemological position that differentiates the social cognitive theory from the cognitive theory. Reciprocal determinism is a feedback causal model of the relationship between behavioral factors, personal factors, and environmental factors (Bandura, 2012). That is, an individual's behavioral changes are consistently being regulated and modified by interacting with social factors in the environment whose feedback influences the next actions and outcomes.

Therefore, it is argued that perceived self-efficacy being an integral part of the personal factors cannot be a fixed trait. It changes in response to changes that occur to the rest of the factors in the reciprocal deterministic system (Bandura, 2012). As it concerns mathematics learning, Borgonovi and Pokropek (2019) conceptualized and described reciprocal determinism as "the sets of relationships underlying the interactions between (a) individuals' exposure to mathematics tasks, (b) mathematics self-efficacy beliefs, and (c) mathematics ability" (p. 269). Therefore, it follows logically to argue that mathematics perceived self-efficacy (henceforth, self-efficacy) is a task-specific construct and affects the performance of engineering students in calculus tasks. Earlier studies have investigated the task-specificity of self-efficacy and confirm that proper attention to task-specificity is a satisfactory way to improve the predictive power of self-efficacy on students' performance in mathematics (Pajares and Miller, 1995). In the present study, the implications of the task-specificity of self-efficacy go beyond the prediction of performance but extend to the research focus and adoption of a self-efficacy measure whose detail is presented in the "Materials and Methods" section.

The concept of self-efficacy has emerged from the social cognitive theory to become a theory on its own. According to the self-efficacy theory, there are four primary sources of self-efficacy beliefs: *enactive mastery experience,* i.e., personal previous task-based achievement, *vicarious experience*, i.e., experience gained by monitoring peers or people around, *verbal/social persuasions*, i.e., complementary or contradictory feedback received from others, and *physiological or affective states,* i.e., physical or emotional situations during the behavioral changes (Bandura, 2008). Among the sources of influence of self-efficacy, previous task-based achievement has been shown empirically to have the most significant impact on students' self-efficacy on mathematics tasks (e.g., Joët et al., 2011; Zientek et al., 2019). Further, Yurt (2014) showed that, apart from predicting self-efficacy, mastery experience has a highly significant correlation with students' mathematics achievement as measured by the end of the semester course grades. As such, if pre-university mathematics content knowledge is considered to be part of the personal previous task-based achievement, then a causal effect is expected between prior mathematics knowledge and the self-efficacy of engineering students. Therefore, the following hypothesis is formulated:

> *Hypothesis one:* There is a direct effect of prior mathematics knowledge on self-efficacy among first-year engineering students.

Fundamental goals of self-efficacy theory within the teaching and learning context are to explain, predict and evaluate differences in students' performance that are brought about by their self-efficacy (Bandura, 2012). A high sense of self-efficacy instills confidence on students' minds when confronted with difficult and challenging mathematical tasks and as such, enables the students to persevere, so that desired outcomes are achieved. In contrast, students with a low sense of self-efficacy cannot forebear difficult situations, doubt their ability, and as such, perform poorly on the learning material. Roick and Ringeisen (2018) reported a longitudinal study in which the contribution of self-efficacy to students' performance in mathematics was investigated. They used a structural equation modelling (SEM) approach with a sample of 206 university students and found that self-efficacy predicts students' performance. Similar corroborative findings on the predictive power of self-efficacy as it concerns university mathematics can be found, elsewhere (e.g., Pajares and Miller, 1994; Pajares and Kranzler, 1995). However, as it is highlighted in the introduction section of the present article, some of these studies have one limitation or the other that makes it difficult to deduce substantial causal claims between self-efficacy and students' performance in mathematics. More so, it could be erroneous to assume that findings from other fields generalize to the university mathematics context considering the task-specificity of self-efficacy. Instead, the author draws on these studies and some postulates of self-efficacy theory to formulate the following hypotheses:

> *Hypothesis two*: There is a direct effect of self-efficacy on engineering students' performance in a first-year calculus course.
> *Hypothesis three*: Self-efficacy mediates the effect of engineering students' prior mathematics knowledge on their performance in a first-year calculus course.

MATERIALS AND METHODS

Research Focus

The present study focuses on the engineering students following a first-year mathematics course at a Norwegian university. Students enrolled in a first-year mathematics course are chosen as participants in the present study for several reasons. First, the author can assess their pre-university mathematics knowledge effectively better than that of students in year two, year three and year four. Second, they are more susceptible to poor performance, high anxiety, and lack of confidence due to their transition from secondary school to university and newness to the university culture. In line with the task-specificity of self-efficacy, data collected from students enrolled on a common mathematics course are more likely to be objective and when analyzed could give a close estimation of the causal relationship between the research constructs. Further, engineering students are the target group in the present study because they form the largest student population following a common mathematics course in the university.

Sample of the Study

An effective sample of 189 engineering students voluntarily participated in the study, most of whom are men (75%). Their age distributions are as follows: 17–20 years (31%), 21–25 years (49%), 26–35 years (15%), and over 36 years (5%). The inclusion and exclusion criteria are based on voluntary consent. As such, the sample can be characterized as a convenient sample. The language of instruction in the course is Norwegian as well as the language used for the mandatory exercises and examinations.

Measures

Prior Mathematics Knowledge

The author adopted a Norwegian mathematics test as a proxy to expose the prior mathematics content knowledge of the participating students in the present study. The test was designed by the Norwegian Mathematical Council to assess pre-university mathematics content knowledge, and it is administered every two years, independent of the present study, to first-year students across several universities and colleges in Norway. It is a 22-item test in which questions are formulated based on the secondary school curriculum. It is assumed that the test is most appropriate in the present study because it has been developed within the Norwegian context and consistently been applied to serve a similar purpose as that of the present study, for the past three decades. Further, the construct validity and the reliability index (using Omega coefficient) of the test have been investigated using a latent variable approach in Mplus 8.3 program, and the latter was found to be 0.92 which together with the unidimensionality of the test show high internal consistency of its items (Zakariya et al., 2020a). However, only a portion of the test (17 items, henceforth, PKMT – prior knowledge of mathematics test) that is of high psychometric properties such as appropriate item difficulty indices (−2.795 to 0.923), item discrimination indices (0.421–1.354), item reliability (0.151–0.646), and unidimensionality, i.e., all the 17 items expose a common latent construct (Zakariya et al., 2020a), is used in the present study. The 17-item PKMT has only two standard multiple-choice questions, and the remaining 15 questions require short answers. All the questions examine the basic knowledge of operations with fractions, decimals, percentages, ratios, similar triangles, speed and distance, and some word problems. A score of 1 point was assigned to a correct answer and a 0 point, otherwise.

Calculus Self-Efficacy

Following the task-specificity of the self-efficacy, the calculus self-efficacy inventory (CSEI) was adopted in the present study. The CSEI was developed with a specific purpose of exposing students' self-efficacy in solving some mathematical tasks drawn from the first-year introductory calculus course (Zakariya et al., 2019). According to the self-efficacy theory, such an inventory offers the best precision in exposing the construct (Bandura, 2006). The CSEI has two parts: preliminary and main parts. The preliminary part of the CSEI contains questions on gender, age, and grade points of students in the highest upper secondary school mathematics course (HGP) they followed before their enrollment into the university. Responses of students to the question on HGP, in addition to the PMKT, are used as proxies to measure their prior mathematics content knowledge. The response values on this item ranging from 1 to 6 points depending on the grades. Further, the main part of the CSEI contains 13 items on exam-type mathematics tasks in which the contents are drawn from the current course curriculum followed by the students. The responses of students on this part of CSEI are used as proxies to expose the latent construct of self-efficacy. The students rate their confidence, on a scale of 0–100, in their belief that they can successfully solve the mathematics tasks. The conceptualization, operationalization, and psychometric properties of the CSEI have been previously studied using factor analysis in FACTOR program coupled with Spearman's rank correlation and well documented (Zakariya et al., 2019). The CSEI was found to possess construct and discriminant validity, unidimensionality, and with a reliability index of 0.90 using ordinal coefficient alpha (Zakariya et al., 2019).

Performance

Finally, the current performance of students in the present study is operationalized and measured by their final scores achieved in the first-year introductory calculus course they followed. It is presumed in the present study, and consistent with the literature (e.g., Cano et al., 2018), that such scores offer the best opportunity to compare individual performance in the course.

Data Collection and Ethical Considerations

The data used in the present study are collected mainly through an online platform, SurveyXact. The author together with his research team independently converted the PKMT to an online test after being granted permission to access the test by the Norwegian Mathematical Council. Similarly, an online version of the CSEI was also prepared. The students were informed of the purpose of the study at a class visit before data collection. Their voluntary consent to take part in the study was sought. As such, they were promised of no consequence, whatsoever, for anyone who decides not to participate in the study. The students were informed that their data will be treated with a high level of security and confidentiality in line with the regulations of the Norwegian Centre for Research Data.

The data were collected on three occasions. At the first occasion, the PKMT was administered in which 40 min of class time was used on the test. This test administration took place in the early weeks of the Autumn semester 2019 because the beginning of the semester is the best time to assess pre-university mathematics content knowledge. On the second occasion, toward the last week of lectures in the Autumn semester 2019, the researchers administered the CSEI through students' registered emails with the university. Because items of the CSEI are drawn from the ongoing mathematics course curriculum, the administration of CSEI was deliberately delayed until the end of the semester. This delay was aimed at ensuring a substantial part of the course curriculum had been covered. The collected data from the two occasions were merged to form an effective sample for the study. In order to ensure the personal data protection regulations are met, the students' administrative affairs office was involved in the process when it came to collating identifiable data. The researcher simply sent the generated survey data to the examination office where the individual final examination scores in the course were added. Afterward, the examination office removed any identifiable information from the data set, and the researcher was provided with a completely anonymized data set. This procedure constitutes the third occasion of the data collection. The data were screened for out of range values, missing values, and normal distribution, all of which pose no challenge to the analyses.

Data Analysis

The Hypothesized Model and Choice of an Instrument

The hypothesized model of the relationship between the calculus self-efficacy (CSE), prior mathematics knowledge (HGP and PKMT), and students' performance in the course (Exam) is presented in **Figure 1**. The main aim of evaluating this model is to estimate the effects of CSE and HGP on Exam. However, there is a challenge with the model. This is because there are some omitted variables, such as the similarity between items on the CSEI and the final examination. The omitted variables act as common causes of both the CSE and the Exam, thereby causing the errors e_1 and e_2 to correlate. This correlation may bias the estimate of the effect of self-efficacy on performance, and thereby constitutes an endogeneity problem in the model (Antonakis et al., 2010). CSE is an endogenous variable in the model because both HGP and PKMT predict it, and it predicts Exam. A way to circumvent this problem, so that a reliable estimate of the effect of self-efficacy on performance can be found is to introduce an instrumental variable, simply called an *instrument*, in the model (Greenland, 2000). It is assumed that the omitted variables do not affect both HGP and PKMT because they are exogenous variables, i.e., they are not predicted by any variable in the model, and as such do not need an instrument. The double-headed arrow between HGP and PKMT in **Figure 1** is a standard notation for correlation between the variables in the SEM literature. It should not be confused with a feedback effect.

An *instrument* "I" is an exogenous variable that satisfies the following properties: (a) "I" has a direct effect on the endogenous variable (CSE) that needs an instrument; (b) The direct effect of "I" on the outcome variable (Exam) is close to zero or completely negligible in the presence of the endogenous variable; (c) "I" should not correlate with the errors associated with the outcome variable (Greenland, 2000; Antonakis et al., 2010). The preliminary analysis in the present study shows that PKMT is the only variable that satisfies the properties (a)–(c), and thus, it was selected as an instrument to discern the true effect of self-efficacy on the performance from the omitted causes in the model.

The Procedure of Data Analysis

The collected data are analyzed using the SEM approach to evaluate the model presented in **Figure 1** and as such, to confirm the plausibility of the research hypotheses. The SEM approach was adopted in the present study because it offers the best and most robust modeling capacity to evaluate causal hypotheses (Bollen and Pearl, 2013). SEM does it better than the path analysis, multiple linear regression, and the partial-least square techniques (Antonakis et al., 2010). Because PKMT was dichotomously scored, the weighted least square mean and variance adjusted (WLSMV) estimator was used which has been shown to provide satisfactory parameter estimates in the analysis of categorical data (Suh, 2015). The author ascertains the "data fitness" of the hypothesized model by looking at both global and local fit indices and parameters. The global fit criteria used are chi-square ratio to the degree of freedom of less than 3, comparative fit (CFI) and Tucker-Lewis indices of greater than or close to 0.90 (Bentler, 1990), and a root mean square error of approximation (RMSEA) value of less than 0.08 (Brown, 2015). The local fits of the model parameters are ascertained by looking at the magnitude and the significant levels of factor loadings, standard errors, and the residual variance, in line with the best practice in SEM literature (Marsh et al., 2004). All the analyses were performed in Mplus 8.3 program.

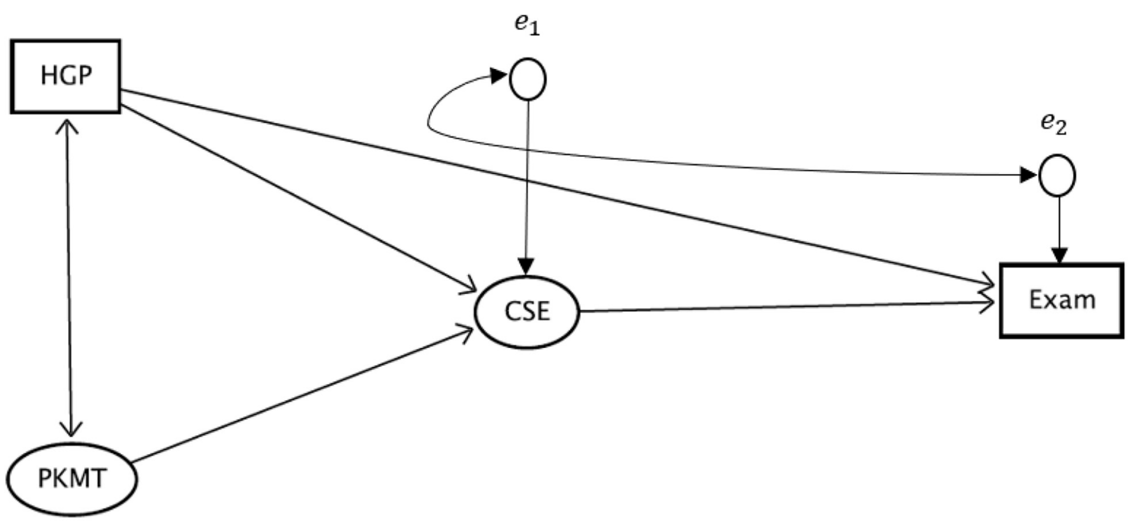


FIGURE 1 | The hypothesized model of the relationship between prior mathematics knowledge, self-efficacy, and students' performance in an introductory calculus course. Both HGP and PKMT are measures of the prior mathematics knowledge of the students, CSE is a measure of the self-efficacy, and Exam represents a measure of performance. The items of both PKMT and CSE are not included in **Figure 1** to enhance the readability of the figure.

RESULTS

The first set of results are from the evaluations of one-factor models for each of the prior mathematics knowledge test and the calculus self-efficacy measurement models. These measurement models are evaluated separately before an evaluation of the hypothesized structural model. In this way, the author could detect and correct any local misspecification in each of the measurement models. This two-step of measurement-before-structural model evaluation has been proven efficient and highly recommended in SEM literature (Byrne, 2012). The dichotomously scored 17 items of the PKMT are hypothesized to expose a common latent factor (prior mathematics knowledge) and tested. All the factor loadings are freely estimated, and the factor variance is fixed to 1 so that the model is identified (Zakariya et al., 2020a). Similarly, the 13 items of the CSEI are hypothesized to expose a common latent factor (self-efficacy) and tested. The factor loadings are freely estimated, the factor variance is fixed to 1, two error covariances between item 09 and item 11 as well as between item 12 and item 13 are allowed in the model as recommended by Zakariya et al. (2019). Further, a maximum likelihood with robust standard errors (MLM) estimator was used instead of the WLSMV because the students' responses on the CSEI are continuous and not categorical. The results from these analyses with regards to the selected global fit indices are presented in **Table 1**.

The results presented in **Table 1** show that the global fit indices are within the recommended ranges for acceptable model fits of the analyzed data. In particular, the ratios of chi-square values to the degrees of freedom, the CFI and the TLI values suggest an acceptable fit for both the PKMT and CSEI models. The RMSEA value and its associated 90% confidence interval with a non-significant *p*-value of the PKMT model show that there is an excellent agreement between the model and the data (Bentler, 1990). Even though the *p*-value of the 90% confidence interval for the RMSEA value in CSEI model is significant, the estimate is lower than 0.08, which suggests a good fit (Brown, 2015). The factor loadings are significant and moderately high, the standard and residual errors are low which are suggestive of acceptable local fit statistics for both the PKMT and CSEI models (Marsh et al., 2004). As such, the author proceeds to the evaluation of the hypothesized structural model, as presented in **Figure 1**, and the resulting global fit indices are presented in **Table 2**. Further, **Figure 2** presents the standardized estimates of the causal effects between the research variables.

The results presented in **Table 2** show an excellent model fit of the evaluated hypothesized structural relationship between the research variables. An excellent model fit in the sense that there is a substantial agreement between the hypothesized model and the analyzed data. This model fit can be deduced from the selected global fit indices that are within the recommended ranges. The ratio of chi-square estimate to the degree of freedom is far less than 3. The CFI and TLI indices are greater 0.95, which indicate an excellent model fit according to the cutoff criteria by Hu and Bentler (1999). The RMSEA estimate together with its perfect (*p*-value = 1.000) 90% confidence interval, suggested that there is a substantial-close fit between the model and analyzed data (Brown, 2015). The global fit indices presented in **Table 2** strengthen the plausibility of the standardized estimates of the causal effects presented in **Figure 2**.

The results presented in **Figure 2** show reliable estimates of the standardized causal effects between the research variables. The reliability of these estimates has been strengthened by the excellent global fit indices reported in **Table 2**. **Figure 2** shows a significant direct effect of PKMT (β = 0.52, standard error – SE = 0.01, $p < 0.001$) on self-efficacy. The direct effect of HGP on self-efficacy is negative and not significant ($\beta = -0.12$, SE = 0.09, $p > 0.05$). Even though, one would have expected a positive effect of HGP on self-efficacy given that students with high grade points in upper secondary school mathematics are expected to have high self-efficacy. The result of the present study does not conform to this expectation. These results show that among the two measures of prior mathematics knowledge, it is only the scores of students on the pre-university mathematics test that have a substantial effect on students' self-efficacy. As such, Hypothesis one is confirmed. The correlation between PKMT and HGP is significant (r = 0.31, SE = 0.08, $p < 0.001$), and it is expected. This is because both PKMT and the HGP are hypothesized to expose different facets of a construct. The correlation between these variables was evaluated instead of a

TABLE 1 | The selected global fit indices for evaluated PKMT and CSEI measurement models.

Global fit indices	PKMT model	CSEI model
Chi-square		
Estimate (χ^2)	143.793	132.162
Degrees of freedom (*df*)	119	64
χ^2/df	1.21	2.065
CFI/TLI		
CFI	0.944	0.911
TLI	0.936	0.892
RMSEA		
Estimate	0.043	0.076
90 percent confidence interval	[<0.001, 0.066]	[0.057, 0.094]
Probability RMSEA ≤ 0.05	0.676	0.013

TABLE 2 | The selected global fit indices of the evaluated hypothesized structural model of the relationship between the research variables.

Global fit indices	Hypothesized model (Figure 1)
Chi-square	
Estimate (χ^2)	492.432
Degrees of freedom (*df*)	458
χ^2/df	1.075
CFI/TLI	
CFI	0.958
TLI	0.954
RMSEA	
Estimate	0.020
90 per cent confidence interval	[<0.001, 0.033]
Probability RMSEA ≤ 0.05	1.000

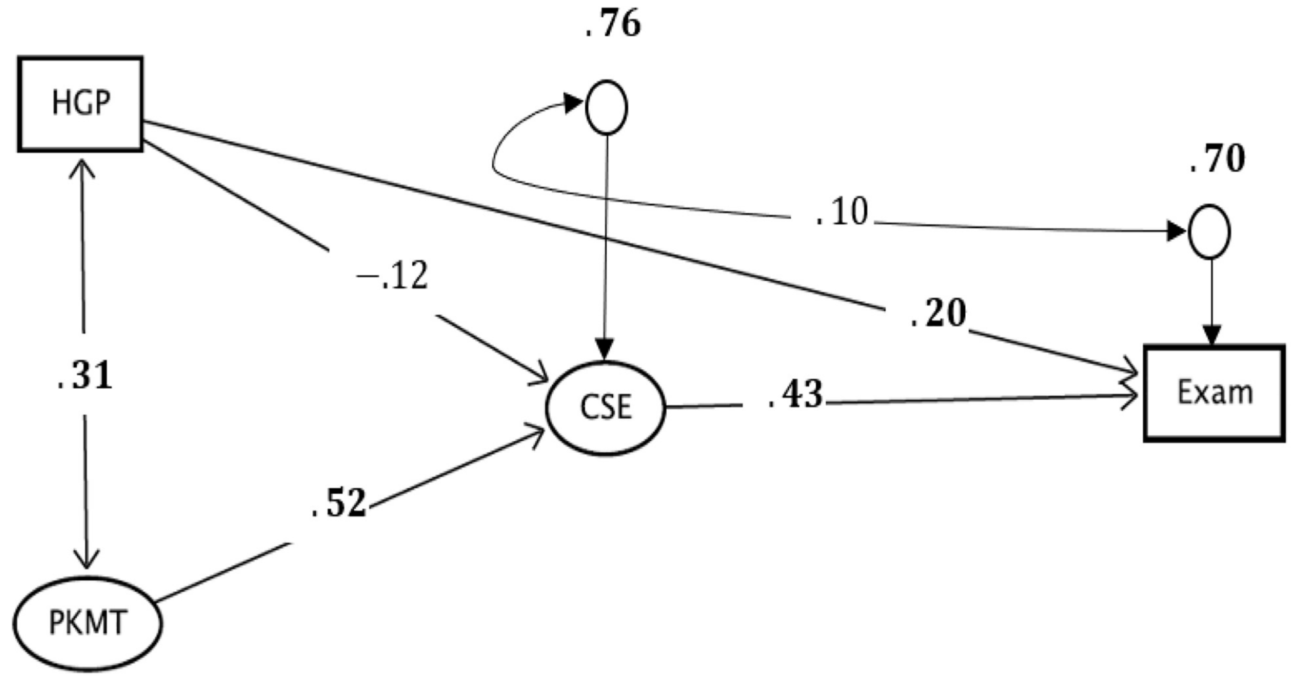


FIGURE 2 | The evaluated model of the relationship between prior mathematics knowledge, self-efficacy, and students' performance in an introductory calculus course. Both HGP and PKMT are measures of the prior mathematics knowledge of the students, CSE is a measure of the self-efficacy, and Exam represents a measure of performance. The significant estimates are in bold faces, and the items of both PKMT and CSE are not included in **Figure 1** to enhance the readability of the figure. The full figure that contains all the items and the associated model parameters is available in **Appendix Figure A1**.

causal relationship for two reasons. The first reason is that they expose different facets of a construct while the second reason is to comply with the recommendations of instrumental variable approach for handling endogeneity problem due to omitted variables in the model (e.g., Kenny, 2012).

It is also revealed in **Figure 2** that the direct effect of self-efficacy on students' performance is significant (β = 0.43, SE = 0.19, p = 0.02) and a significant standardized residual estimate of 0.76. These results confirm the plausibility of Hypothesis two. The residual error shows that the prior mathematics knowledge of students explains 24% of the factor variance in self-efficacy. The percentage of the explained factor variance is moderate, considering the limited number of variables that predict self-efficacy in the model. The error covariance between the self-efficacy and students' performance is not significant (r = 0.10, SE = 0.25, $p > 0.05$) which is a good result as it confirms the reliability of the estimated effect of self-efficacy on performance after introducing the instrument in the model. **Figure 2** also shows that the direct effect of HGP on the students' performance is significant (β = 0.20, SE = 0.07, p = 0.005).

More so, the results of the mediation analysis show the standardized total effect of prior mathematics knowledge (PKMT and HGP) on performance to be 0.37. A significant indirect effect of PKMT through self-efficacy was found (β = 0.22, SE = 0.10, p = 0.03), and a non-significant indirect of HGP on performance through self-efficacy efficacy (β = −0.05, SE = 0.04, $p > 0.05$). These results show that self-efficacy mediates the direct effect of PKMT on performance while that of HGP on performance is not mediated, beyond chances. This finding confirms, in part, the plausibility of Hypothesis three. Finally, the significant standardized residual estimate of 0.70 on the Exam variable in **Figure 2** shows that 30% of the variability in students' performance is explained by both the prior mathematics knowledge and self-efficacy. This variability is considered to be moderately high, and more discussion about this is presented in the next section.

DISCUSSION, LIMITATIONS, AND RECOMMENDATIONS

Discussion of Findings

Self-efficacy has been articulated theoretically to be an important construct in explaining variability in students' performance. Several pieces of empirical evidence have demonstrated its relevance to students' performance in psychology, sport, and clinical medicine (Bandura, 1997). Meanwhile, due to the task-specificity of self-efficacy, it could be erroneous to assume generalization of findings from other fields to the mathematics learning context. More so, there are limited studies with a focus on mathematics self-efficacy and its effects on students' performance in university mathematics. As such, attempts are made in the present study to investigate the causal effects of mathematics self-efficacy on students' performance through an innovative approach of instrumental variable modeling (Greenland, 2000). Prior mathematics knowledge (PKMT and HGP) and self-efficacy (CSEI) are conceptualized and operationalized based on previous studies and the self-efficacy theory. The measurement model of PKMT was evaluated, and it was found to provide reliable estimates of the construct it was hypothesized to expose. The findings of the present study also confirm reliable estimates of the measurement model of CSEI. These findings are consistent with the findings of previous studies on the two measures (Zakariya et al., 2019, 2020a). After establishing acceptable measurement models of the two measures, the hypothesized structural relationship between the research constructs was evaluated. The major findings are discussed in the forthcoming paragraphs.

The results of the present study confirm a direct effect of prior mathematics knowledge test on students' calculus self-efficacy. This finding can be interpreted to mean that students with high scores on the prior mathematics knowledge test have a high sense of self-efficacy in solving first-year calculus tasks successfully. This finding is consistent with the postulated impact of personal previous task-based achievement on self-efficacy by the self-efficacy theory (Bandura, 2012). It was found that prior mathematics knowledge test alone accounts for 27% (i.e., the square of 0.52 times 100%) of the variability of the self-efficacy. However, this percentage of explained variance reduced to 24% when this direct effect of the test scores is combined with the direct effect of HGP on self-efficacy. The direct effect of prior knowledge of mathematics test on self-efficacy found in the present study is far higher than the effects of high school level, and the college credits (both operationalized to measure prior experience) on students' self-efficacy in completing mathematics problem-solving tasks reported, elsewhere (Pajares and Miller, 1994; Pajares and Kranzler, 1995). Given that these studies are relatively old and the mathematics curriculum in higher education is changing to catch up with our 21st-century challenges, it is claimed that the present finding is novel and the captures current situation on the causal relation between prior mathematics knowledge and self-efficacy among university students.

Another major finding of the present study is the exposed direct effect of calculus self-efficacy on students' performance in the course. A unique feature about the estimate of this direct effect lies in the ability of the instrumental variable approach to discern this effect from that of other disturbances which affect students' performance but are not included in the model. This finding is interpreted to mean a high sense of self-efficacy is a potential cause of high scores of students, beyond chances, in the first-year introductory calculus course. By implication, this finding provides empirical support for designing interventions that foster self-efficacy as proxies to enhance students' performance in the first-year introductory mathematics course. Such interventions may be in the inform of realistic modeling of the links between previous achievements and self-efficacy, social persuasion by older students who have passed the course, and other related activities that can be traced to the sources of self-efficacy. The magnitude of the estimated causal effect of self-efficacy on students' performance in the present study is substantially higher than comparable direct effects reported in previous studies (Pajares and Kranzler, 1995; Roick and Ringeisen, 2018). As such, the author claims that the causal relationship exposed between self-efficacy and performance by the findings of the present study has a significant contribution to mathematics education literature.

Apart from the substantial contribution of the calculus self-efficacy to students' performance exposed in the present study, a major finding is the detected mediating role of self-efficacy between prior knowledge mathematics test and students' current performance in the course. It was found in the present study that about 46% (i.e., 0.17 out of 0.37) of the total effect of prior mathematics knowledge (PKMT and HGP) on students' performance is mediated by self-efficacy. On the one hand, this finding may be interpreted to mean students with high scores on both the prior knowledge of mathematics test and the self-efficacy performed, beyond chances, better than the students who do not score high on the two measures. On the other hand, it confirms the mediating role of self-efficacy as postulated by the self-efficacy theory (Bandura, 1997). This finding also corroborates the mediating role of mathematics self-efficacy that is reported, elsewhere, using path analysis (Pajares and Miller, 1994). Despite the limited number of variables the author considered in the evaluated structural model of the relationship between the research constructs, the percentage of the explained variance (30%) in students' performance is higher than the reported values in studies with several predictor variables (Pajares and Miller, 1994, 1995). It is conjectured that the task-specificity of the self-efficacy measure coupled with the innovative instrumental variable approach used in the present study contributes to the moderately high percentage of explained variance in the students' performance. Potential variables that could increase the percentage of explained variance, if included in the model, are approaches to learning mathematics, academic motivation, mathematics anxiety, and attitudes toward mathematics learning. Future studies are recommended with this intention.

Potential Limitations and Recommendations

A potential limitation of the present study is attributable to the restriction of sample to first-year engineering students enrolled on a course. Even though this restriction offers several advantages as previously highlighted in the "Materials and Methods" section, it might also hinder the generalization of the findings beyond a similar student population. Future replicated studies are recommended with a focus on students following a variety of courses at different levels of higher education. However, such studies should devise innovative ways or use robust statistical modeling such as multi-level SEM combined with the instrumental variable approach to account for task-specificity of the self-efficacy across diverse populations. Also, the relatively small sample size (189 students) could be a threat to the validity of the SEM results given that some researchers have recommended higher sample sizes (Marsh et al., 1998; Byrne, 2012). However, it has been theoretically argued and empirically shown that a "one size fits all" rule is not tenable for sample sizes of SEM studies (Wolf et al., 2013). As such, sample sizes close to 200 cases are recommended for conducting SEM studies that involve moderately complex models (Kline, 2016). Notwithstanding, future replication studies are recommended with a larger sample size to cross-validate the findings of the present study.

More so, the self-efficacy theory postulates a feedback causal relationship between self-efficacy and students' performance in mathematics through reciprocal determinism model (Borgonovi and Pokropek, 2019). Nevertheless, the focus of the present study is only on one-directional causal effect from self-efficacy to students' performance which could also constitute a limitation. The author argues that such a feedback causal relationship is better investigated using a longitudinal research design (e.g., Roick and Ringeisen, 2018) than the survey research design used in the present study. As such, future longitudinal studies are recommended with this intention. The author also acknowledges that a limited number of predictor variables in the evaluated structural model of the present study may constitute another limitation. Had been more relevant variables such as approaches to learning, motivation, and mental ability that have been linked with performance are included in the model (Pajares and Miller, 1994; Zakariya et al., 2020b), the percentage of explained variance in students' performance would have improved. Future study may also be conducted with this intention.

CONCLUSION

The present study is motivated by the lack of empirical evidence on the causal relationship between self-efficacy and students' previous and current performance in university mathematics. Therein, attempts are made to fill this gap by investigating hypothesized causal claims between the research constructs using the instrumental variable approach to modeling. The major findings in the present study establish a causal relationship with reliable estimates between self-efficacy and students' performance in an introductory calculus course at a university in Norway. The author conjectures that these findings are generalizable to similar student populations within and beyond Norwegian borders. This conjecture is based on both theoretical and innovative statistical perspectives adopted in the present study. As such, the author recommends replication of the present study to investigate this conjecture within the quantitative research paradigm. The author declares that an outright discovery of the causal relationship between self-efficacy and students' performance in mathematics is not claimed in the present study. Instead, it is hoped that foundations are laid for future experimental, randomized-control trial studies with this intention.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Norwegian Centre for Research Data. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YZ: conceptualization, methodology, formal analysis, software, data curation, investigation, visualization, and writing-original draft preparation.

ACKNOWLEDGMENTS

The author acknowledges the support of Faculty of Engineering and Science, University of Agder, Kristiansand, Norway as well as MatRIC, Centre for Research, Innovation and Coordination of Mathematics Teaching. Also, S. Goodchild, K. Bjørkestøl, and H. K. Nilsen are acknowledged for their useful comments on the first draft of the manuscript.

REFERENCES

Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2010). On making causal claims: a review and recommendations. *Leadership Q.* 21, 1086–1120. doi: 10.1016/j.leaqua.2010.10.010

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.

Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Ann. Rev. Psychol.* 52, 1–26. doi: 10.1146/annurev.psych.52.1.1

Bandura, A. (2006). "Guide for constructing self-efficacy scales," in *Self-Efficacy Beliefs of Adolescents*, Vol. 5, eds F. Pajares and T. Urdan (Greenwich, CT: Information Age Publishing), 307–337.

Bandura, A. (2008). "An agentic perspective on positive psychology," in *Praeger perspectives. Positive psychology: Exploring the best in people*, ed. S. J. Lopez (Westport, US: Praeger Publishers/Greenwood Publishing Group), 167–196.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *J. Manage.* 38, 9–44. doi: 10.1177/0149206311410606

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238

Bollen, K. A. (2019). Model implied instrumental variables (MIIVs): an alternative orientation to structural equation modeling. *Multivariate Behav. Res.* 54, 31–46. doi: 10.1080/00273171.2018.1483224

Bollen, K. A., and Pearl, J. (2013). "Eight myths about causality and structural equation," in *Handbooks of Sociology and Social Research*, ed. S. L. Morgan (Dordrecht: Springer), 301–328.

Borgonovi, F., and Pokropek, A. (2019). Seeing is believing: task-exposure specificity and the development of mathematics self-efficacy evaluations. *J. Educ. Psychol.* 111, 268–283. doi: 10.1037/edu0000280

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, London: The Guilford Press.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge, Taylor & Francis Group.

Cano, F., Martin, A. J., Ginns, P., and Berbén, A. B. G. (2018). Students' self-worth protection and approaches to learning in higher education: predictors and consequences. *Higher Educ.* 76, 163–181. doi: 10.1007/s10734-017-0215-0

Czocher, J. A., Melhuish, K., and Kandasamy, S. S. (2019). Building mathematics self-efficacy of STEM undergraduates through mathematical modelling. *Int. J. Math. Educ. Sci. Tech.* 51, 1–28. doi: 10.1080/0020739x.2019.1634223

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.* 29, 722–729. doi: 10.1093/ije/29.4.722

Hackett, G., and Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance. *J. Res. Math. Educ.* 20, 261–273.

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struc. Eq. Model. Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118

Joët, G., Usher, E. L., and Bressoux, P. (2011). Sources of self-efficacy: an investigation of elementary school students in france. *J. Educ. Psychol.* 103, 649–663. doi: 10.1037/a0024048

Kenny, D. A. (2012). *Estimation with instrumental variables*. Available Online at: http://davidakenny.net/cm/iv.htm

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Lahdenperä, J., Postareff, L., and Rämö, J. (2019). Supporting quality of learning in university mathematics: a comparison of two instructional designs. *Int. J. Res. Undergrad. Math. Educ.* 5, 75–96. doi: 10.1007/s40753-018-0080-y

Marsh, H. W., Hau, K.-T., Balla, J. R., and Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behav. Res.* 33, 181–220. doi: 10.1207/s15327906mbr3302_1

Marsh, H. W., Hau, K.-T., and Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struc. Eq. Model. Multidiscip. J.* 11, 320–341. doi: 10.1207/s15328007sem1103_2

Özcan, Z. Ç, and Eren Gümüş, A. (2019). A modeling study to explain mathematical problem-solving performance through metacognition, self-efficacy, motivation, and anxiety. *Aus. J. Educ.* 63, 116–134. doi: 10.1177/0004944119840073

Pajares, F., and Kranzler, J. H. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemp. Educ. Psychol.* 20, 426–443. doi: 10.1006/ceps.1995.1029

Pajares, F., and Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *J. Educ. Psychol.* 86, 193–203. doi: 10.1037/0022-0663.86.2.193

Pajares, F., and Miller, M. J. (1995). Mathematics self-efficacy and mathematics performances: the need for specificity of assessment. *J. Couns. Psychol.* 42, 190–198. doi: 10.1037/0022-0167.42.2.190

Peters, M. L. (2013). Examining the relationships among classroom climate, self-efficacy, and achievement in undergraduate mathematics: a multi-level analysis. *Int. J. Sci. Math. Educ.* 11, 459–480.

Roick, J., and Ringeisen, T. (2018). Students' math performance in higher education: Examining the role of self-regulated learning and self-efficacy. *Learn. Indiv. Differ.* 65, 148–158. doi: 10.1016/j.lindif.2018.05.018

Rozgonjuk, D., Kraav, T., Mikkor, K., Orav-Puurand, K., and Täht, K. (2020). Mathematics anxiety among STEM and social sciences students: the roles of mathematics self-efficacy, and deep and surface approach to learning. *Int. J. STEM Educ.* 7:46. doi: 10.1186/s40594-020-00246-z

Schukajlow, S., Achmetli, K., and Rakoczy, K. (2019). Does constructing multiple solutions for real-world problems affect self-efficacy? *Educ. Stud. Math.* 100, 43–60. doi: 10.1007/s10649-018-9847-y

Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Struc. Eq. Model. Multidiscip. J.* 22, 568–580. doi: 10.1080/10705511.2014.937669

Wolf, E. J., Harrington, K. M., Clark, S. L., and Miller, M. W. (2013). Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educ. Psychol. Meas.* 76, 913–934. doi: 10.1177/0013164413495237

Yurt, E. (2014).). The predictive power of self-efficacy sources for mathematics achievement. *Educ. Sci.* 39, 159–169. doi: 10.15390/eb.2014.3443

Zakariya, Y. F., Goodchild, S., Bjørkestøl, K., and Nilsen, H. K. (2019). Calculus self-efficacy inventory: Its development and relationship with approaches to learning. *Educ. Sci.* 9:170. doi: 10.3390/educsci9030170

Zakariya, Y. F., Nilsen, H. K., Goodchild, S., and Bjørkestøl, K. (2020a). Assessing first-year engineering students' pre-university mathematics knowledge: Preliminary validity results based on an item response theory model. *J. Technol. Sci. Educ.* 10, 259–270. doi: 10.3926/jotse.1017

Zakariya, Y. F., Nilsen, H. K., Goodchild, S., and Bjørkestøl, K. (2020b). Self-efficacy and approaches to learning mathematics among engineering students: empirical evidence for potential causal relations. *Int. J. Math. Educ. Sci. Technol.* 2020, 1–15. doi: 10.1080/0020739X.2020.1783006

Zientek, L. R., Fong, C. J., and Phelps, J. M. (2019). Sources of self-efficacy of community college students enrolled in developmental mathematics. *J. Further Higher Educ.* 43, 183–200. doi: 10.1080/0309877x.2017.1357071

Zuffianò, A., Alessandri, G., Gerbino, M., Kanacri, B. P. L., Di Giunta, L., Milioni, M., et al. (2013). Academic achievement: the unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learn. Indiv. Differ.* 23, 158–162. doi: 10.1016/j.lindif.2012.07.010

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

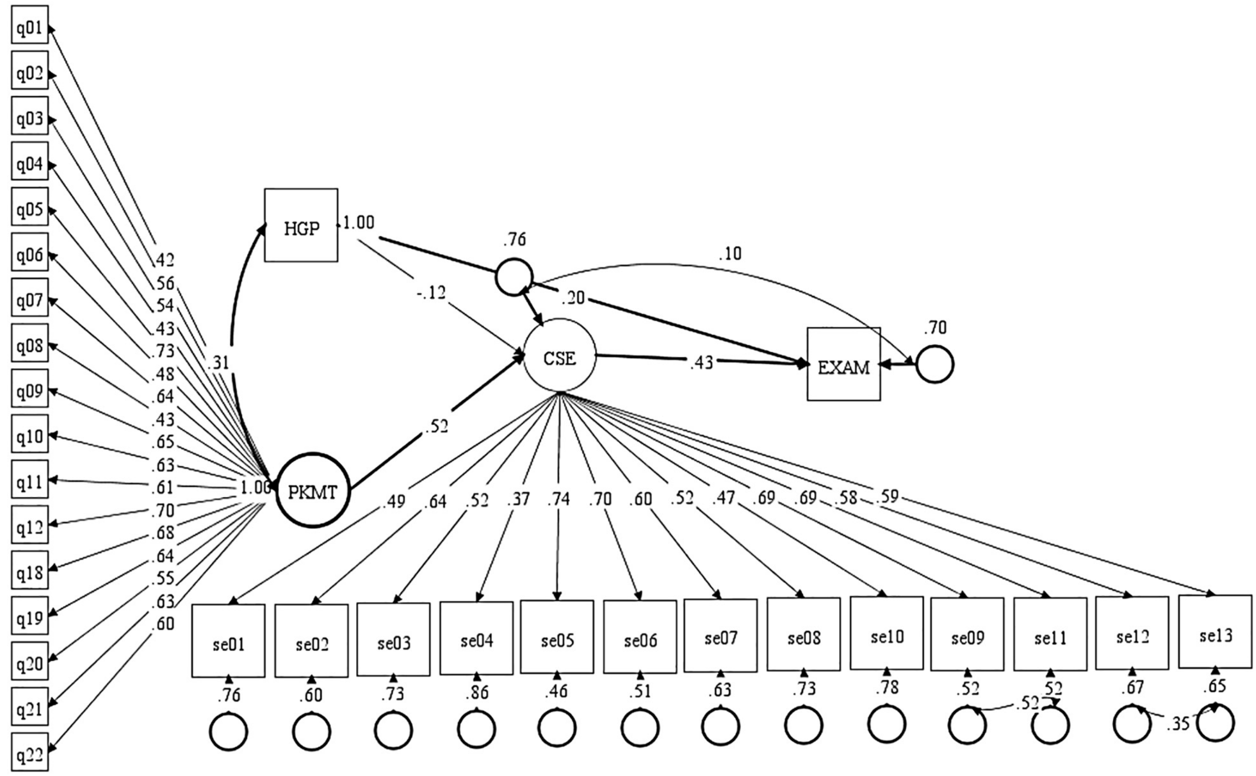


FIGURE A1 | The full evaluated model of the relationship between prior mathematics knowledge, self-efficacy, and students' performance in an introductory calculus course. The significant paths are in bold faces.