Assessment of citizenship competences

Daas, R.J.M.

[Link to publication](#)

ASSESSMENT OF CITIZENSHIP COMPETENCES

Remmert Daas

ASSESSMENT OF CITIZENSHIP COMPETENCES

Remmert Daas

ASSESSMENT OF CITIZENSHIP COMPETENCES

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op woensdag 3 april 2019, te 14:00 uur

door

Reinbert Johannes Maria Daas

geboren te Hoorn

Promotiecommissie

Promotor(es):   Prof. dr. A.B. Dijkstra, Universiteit van Amsterdam

                Prof. dr. S. Karsten, Universiteit van Amsterdam

Overige leden: Dr. J.G. Janmaat, University College London

                Prof. dr. G.J. Overbeek, Universiteit van Amsterdam

                Prof. dr. W.M.M.H. Veugelers, Universiteit voor Humanistiek

                Prof. dr. M.L.L. Volman, Universiteit van Amsterdam

                Prof. dr. H.G. van de Werfhorst, Universiteit van Amsterdam

Faculteit der Maatschappij- en Gedragswetenschappen

CONTENTS

CHAPTER 1: INTRODUCTION

Education is widely considered to contribute to the development of young people to participate in society. Over the past decades many countries have introduced legislation mandating schools to promote the development of young people into actively engaged citizens (Eurydice, 2017). The attention for citizenship education appears to stem from a dissatisfaction with (young) citizens' political participation and lack of social cohesion in society (e.g. Biesta, 2009; Bron, 2006; Crick, 2007). Citizenship education aims to – *inter alia* – promote young people's political attitudes, understanding the workings of democratic institutions, ability to take part in democratic decision-making, and views on social issues (Council of Europe, 2010; Eurydice, 2017; Galston, 2001; Veugelers, 2007).

The premise of citizenship education is that these competences do not develop naturally, and that specific efforts by schools, teachers and students are necessary. However, insights into how students (best) develop these competences is inconclusive; in no small part because of the difficulties in assessing citizenship competences (Kerr, Keating, & Ireland, 2009). This introduction discusses three questions on the assessment of citizenship competences that are the starting point to this dissertation: *What* are citizenship competences? *Why* should we want to assess citizenship competences? and *How* can citizenship competences be assessed?

Citizenship competences

Citizenship can be considered to refer both to citizens' political and social relations within society, as well as social relations between citizens (Crick, 2007; Ten Dam & Volman, 2007). Citizenship refers both to citizens' legal rights and obligations, as well as their willingness and ability to participate in and contribute to their (public) community (Van Gunsteren, 1998). Citizenship competences refer to the capacity to act in these situations (cf. Rychen & Salganik, 2003). Citizenship competences constitute the knowledge, attitudes and skills to act effectively in a plethora of social contexts. More specifically, citizenship competences refer to such concepts as knowledge on democratic representation, trust in state and civil institutions, and intentions to vote, but also to knowledge on cultural differences, attitudes towards democratic values, and skills to take part in democratic debate (cf. Schulz et al., 2016). Citizenship competences in other words refer to the knowledge, attitudes and skills (young) people need to actively participate as members of a democratic society.

Adolescence is considered an important period for the development of these competences (Amnå et al., 2009; Eckstein, Noack, & Gniewosz, 2012; Hooghe & Wilkenfeld, 2008). Schools can promote the development of students' citizenship competences through *inter alia* specifically directed lessons, cross-curricular activities, and democratic school climate (Geboers et al., 2013). Studies of citizenship competences commonly use surveys consisting of tests and questionnaires to measure students' citizenship competences (e.g. Keating et al., 2010; Ten Dam et al., 2011; Schulz et al., 2010, 2017).

Purpose of assessment

Assessment is common practice in education. Students' knowledge, skills and attitudes are assessed throughout their educational career, and these assessments serve a range of functions, e.g.: produce a pass/fail grade, provide diagnostic information, motivate student learning, facilitate school improvement, or for accountability purposes. These functions of assessment are commonly put in two categories: assessment for learning, and assessment of learning (Wiliam & Black, 1996).

Assessment for learning (or: formative assessment) emphasizes that assessment serves to support the learning process. Assessment should inform teachers and students where students are in their learning process, where they need to go, and how (best) to move forward in order to realize their learning aims (Broadfoot et al., 2002). The emphasis on assessment to inform and support future learning means feedback plays an important role here (cf. Hattie, 2008). Learning might also take place at the level of teachers and schools, in which case assessment aims to e.g. facilitate school improvement (cf. Dijkstra & de la Motte, 2014).

Assessment of learning (or: summative assessment) generally aims to measure students achievement, either compared relative to their peers or to set standards. Students are for example given a pass or fail grade based on their performance. Assessment of learning can serve multiple functions. At the student level we might for example be interested to evaluate which students have achieved a predetermined level of proficiency, or which students perform better or worse than their peers. At the school level we might be interested in evaluating if the efforts of schools and teachers have led to the desired outcomes for a group of students. National policies on education differ with between countries and over time, and so at the level of the school system we might be interested to evaluate the effectiveness of education policies (cf. Dijkstra & de la Motte, 2014). Depending

on the purpose of assessment, different assessment approaches may be most suitable – which brings us to our next topic: different modes of assessment.

Different modes of assessment

Student assessment can be done in many different ways. However, assessment in education predominantly consists of written tests. This is also the case for the assessment of citizenship knowledge, where most instruments consist of multiple-choice questions, sometimes supplemented with (short answer) open questions (cf. Keating et al., 2010, Schulz et al., 2010, 2017; Ten Dam et al., 2011). In line with many surveys among adult populations (e.g. World Values Survey, European Social Survey) attitudes and skills are commonly assessed using Likert-type questions, the scores on which are combined to form meaningful scales (ibid.). Several studies have proposed citizenship competences should (also) be assessed in more open-ended approaches such as portfolios, that allow for students to express their personal experiences (Jerome, 2008; Kerr et al., 2009). However, almost ten years after these publications, multiple-choice tests and questionnaires are still most prevalent. This begs the question what methods for assessment of citizenship competences are viable, and new instruments could be developed.

Organization of this dissertation

This dissertation aims to contribute to promote further understanding and development of the assessment of citizenship competences. To this end we investigate the application of different instruments to assess citizenship competences, their strengths and weaknesses, possible implications, and consequences for assessing citizenship competences.

The International Civic and Citizenship Education Study (ICCS) is the largest study of students' citizenship competences across nations. One of the aims of the study is to improve countries' understanding of students' citizenship competences and emerging civic-related challenges (Schulz et al., 2017). Most countries partaking in the study have developed policies on citizenship education in schools (Eurydice, 2017). The developments of mandating citizenship education for schools show strong parallels between European countries. However, countries also differ in the aims and practices of citizenship education (Hahn, 1999). Chapter 2 considers the link between national citizenship education policy and outcomes based on assessment using tests and questionnaires (specifically ICCS 2009). We describe the policies on citizenship education in four countries: the Netherlands,

Norway, Scotland, and Sweden. These countries all introduced policy on citizenship education in recent decades, but each with distinct aims and characteristics. The chapter evaluates whether the attention to issues in policy and debate are reflected in comparative assessment of students' citizenship competences.

The prevalence of tests and questionnaires in the assessment of citizenship competences begs the question if other approaches are also feasible. The next two chapters focus specifically on a less common approach to citizenship competences: rubrics. Rubrics specify a set of criteria, which are described at several levels of performance. Rubrics are considered to promote learning by making expectations and criteria explicit and by facilitating feedback and self-assessment (Andrade, 2005). Chapter 3 asks to what extent rubrics can support the assessment of students' citizenship competences. We developed three rubrics for the assessment of citizenship competences (see Appendix A), and tested these among students in grade 10 general secondary education and grade 11 tertiary vocational education. The results show rubrics can be a viable instrument for the assessment of citizenship competences, but to different extents for citizenship knowledge, attitudes and skills. Rubrics show particular promise for the assessment of students' citizenship attitudes. Chapter 4 therefore focuses more specifically on the implications of using rubrics to support the assessment of citizenship attitudes.

Chapter 5 offers a set of criteria by which to evaluate assessment instruments. These criteria are then applied to evaluate the strengths and weaknesses of tests and questionnaires, portfolio assessment, game-based assessment, and vignettes to assess citizenship competences. Finally, Chapter 6 provides a synthesis of the main findings, the conceptual contributions of this dissertation, its limitations, and implications for furthering the assessment of citizenship competences.

CHAPTER 2: LINKING CITIZENSHIP EDUCATION POLICY TO STUDENTS' CITIZENSHIP COMPETENCES IN THE NETHERLANDS, NORWAY, SCOTLAND AND SWEDEN[1]

Education not only serves to promote cognitive abilities but also has a role in promoting students' social and civic competences. While social competences are often considered in conjunction with school climate, citizenship competence poses an interesting case since it exposes the link between current behavior (at school) and preparation for future engagement with society. Considering the importance of social and civic competence for both individual and society, the extent to which schools have taken up this responsibility expresses an indication of the extent to which schools are able to fulfil this function. The extent to which schools have taken up this role appears to differ both between and within countries (Eurydice, 2012). Comparative studies into students' citizenship competence provide valuable insight into the school's contribution to these developments (see Schulz et al., 2010; Torney-Purta et al., 2001).

This chapter aims to provide insights into educational practice by combining an overview of citizenship education in practice with data on students in lower-secondary education's citizenship competences. To this end, this chapter provides both a discussion of the prevalent aspects of citizenship education in each country, combined with a brief account of the developments that have taken place.

The analyses employed here use data from the 2009 International Civic and Citizenship Education Study (ICCS; Schulz et al., 2010). This study assessed around 3,000 students in each of 38 countries on a range of citizenship aspects, including civic knowledge, attitudes and behavior.[2] These measures were allocated to 25 scales relating to students' citizenship competences. In this chapter we use five selected scales: citizenship knowledge, support for democratic values, attitudes

---

[1] Based on: Daas, R. (2014). Linking citizenship education policy to students' citizenship competences in the Netherlands, Norway, Scotland and Sweden. In A. B. Dijkstra, & P. I. de la Motte (Eds.), *Social outcomes of education: The assessment of social outcomes and school improvement through school inspections* (pp. 73–99). Amsterdam, Netherlands: Amsterdam University Press.
[2] The final set included 1,964 students in the Netherlands, 3,013 in Norway, and 3,464 in Sweden. While the response rate in the Netherlands failed to meet set standards, the results are still considered to be representative (Maslowski et al., 2012). Scotland did not partake in the ICCS 2009 study.

toward equal rights for all ethnic/racial groups, expected adult electoral participation, and civic participation at school.

This chapter considers the features of citizenship education in the Netherlands, Norway, Scotland and Sweden, and links these to student citizenship competences (except in the case of Scotland). Student outcomes are *italicized* in the country sections.[3] In the final section we consider some cross-national issues including school effect size and the implications of different approaches taken to measurement and assessment.

Citizenship education in the Netherlands

Citizenship became a statutory part of education in the Netherlands when in 2006 a law took effect whereby primary and secondary schools in the Netherlands became lawfully obligated to provide schooling in active citizenship and social integration. Active citizenship refers to the readiness and ability to make an active contribution to the community. Social integration refers to participation in society and its institutions regardless of ethnic or cultural background and familiarity with Dutch culture (Ministerie van Onderwijs, Cultuur & Wetenschap, 2005). This explicit link between citizenship and integration has had consequences for the implementation of citizenship in the curriculum. Implementation is largely considered a school responsibility. School autonomy in the Netherlands is the highest of all OECD countries, with around 85 percent of decisions taken at the school level (OECD, 2012).

Citizenship education policy
Prior to its lawful footing, the Education Council defined citizenship to be composed of two aspects; 1) what citizens may and must do (the formal political-juridical side of citizenship), and 2) what citizens can and want to do (the social side of citizenship). The report focuses mostly on the second aspect by promoting the ability and willingness to participate in and contribute to society (Onderwijsraad, 2003). This distinction typically reflects much of the writing done on citizenship in the Netherlands. Much focus is placed on the social aspects of citizenship whereas political content appears to focus mostly on democratic values (see Inspectie van het Onderwijs, 2006). Citizenship is presented in conjunction

---

[3] Average student performance is graded 'low', 'average' or 'high' compared to whether there is a significant difference with cross-national ICCS average.

with stimulating social cohesion and integration (Ministerie van Onderwijs, Cultuur & Wetenschap, 2005).

Citizenship education is not offered as a school subject but is cross-curricular. Core objectives are set for primary and lower-secondary school, several of which are related to citizenship (Ministerie van Onderwijs, Cultuur & Wetenschap, 2006a, 2006b; see also Chapter 5). Bron (2006) concludes that these core objectives and the overall (policy) aims stated for citizenship education only partially connect. Though the core objectives provide a basis for active citizenship, they focus on knowledge and skills, while citizenship education aims are largely based on values and democratic principles. Terms such as willingness and participation are not reflected in the core objectives, while these are central to the aims set for citizenship education. Furthermore, core objectives relating to providing an active contribution to society are lacking.

While the line from policy to school objectives shows discrepancies, there are further indications that the continued line for (envisioned) practice is also flawed. Both primary and secondary school boards view citizenship education as a primary task (Bronneman-Helmers & Zeijl, 2008). However, while they show substantial support for teaching social skills and basic democratic values, attention for democracy and other cultures is much less prominent (Inspectie van het Onderwijs, 2010, 2013). Furthermore, most schools are found not to operate from a planned approach but provision often involves patchwork (Bron & Thijs, 2010, 2011; Inspectie van het Onderwijs, 2011). Though the majority of schools report working on the development of their citizenship education, the development of school practice shows little progress (Inspectie van het Onderwijs, 2010, 2011; Peschar et al., 2010). Classroom practice appears generally non-democratic (Sandstrom Kjellin & Stier, 2008), and teachers are critical of the attainability of the social competence goals associated with citizenship education (Leenders et al., 2008; Zwaans et al., 2006). Finally, pupils' citizenship knowledge at the end of primary school is found to be unsatisfactory (Wagenaar et al., 2011).

In sum, it appears that the line from citizenship education policy to teaching practice and student competence shows large discrepancies. While this lack of alignment can be expected to be reflected in Dutch students' score on civic knowledge, *Dutch students score close to the international average on citizenship knowledge.* It seems a lack of clarity on the subject does not lead schools to refrain from spending time on citizenship education. Nonetheless, Dutch students are outperformed by students in most other European countries.

Themes in citizenship education

Besides the discussions on citizenship education policy, a number of themes can be identified which can be considered to characterize citizenship education in the Netherlands. We briefly discuss some of the most prevailing and typical of these.

Fundamental democratic values

Schools are expected to adhere to and promote seven basic democratic values which support peaceful cohabitation and citizenship: freedom of expression, equality, understanding of others, tolerance, autonomy, rejecting intolerance and rejecting discrimination (Inspectie van het Onderwijs, 2006). Citizenship is based on the support of these democratic principles. Though these values present a reasonable core on which to found citizenship, they allow for a wide range of elaborations. Furthermore, as argued above, there is reason to believe that provision by schools is prone to circumstance. While these basic aims will generally be supported, little explicit teaching appears to be devoted to them. Indeed, *Dutch students score low on support for democratic values.* It seems that students are disengaged with this topic.

Equal rights and integration

Citizenship education in the Netherlands focuses largely on social cohesion and integration. The debate on immigration can be considered prominently visible both in and outside the political arena (Doppen, 2010). Some researchers have noted that integration as envisioned in the Dutch debate expects a one-sided effort from the side of immigrants, compelling non-Western immigrants to assimilate (Leeman & Pels, 2006; Stolz, 2011). This notion can be considered to convey some sense of suspicion towards minority cultures. Schools appear to spend less time on teaching about other cultures than other aspects of citizenship (Inspectie van het Onderwijs, 2013). However, both teachers and schools appear to generally attach more importance to the development of attitudes than other aspects of citizenship (Inspectie van het Onderwijs, 2010, 2013). While the image does show some problems in terms of 'how equal' equality should be, there appears to be much attention for the development of these attitudes. The ICCS data however, show *Dutch students score low on support for equal rights for ethnic/racial groups*. It would appear that the significance attached to this topic by schools and teachers is not (yet) shared by Dutch 14-year-olds.

Participation

Some researchers argue that citizenship education is at risk of promoting future citizenship, where it should regard children as young citizens (see Lawy & Biesta, 2006). Emphasizing current citizenship behavior as opposed to preparing students for future citizenship has implications for envisioned practice. Schools that actively promote citizenship appear to favor an emphasis on the development of students' social skills, with schools acting as a place of practice (Hilbers et al., 2010). When schools are successful at adopting this approach, we can expect positive results on students' citizenship competences. The expectation is, however, that this approach is adopted only in a small proportion of schools, as they are free to choose their own approach, and citizenship education appears to be low-key for most schools. This notion is reflected in the ICCS scores, as *Dutch students score low on civic participation at school.* In fact, the score is far below the international average, indicating some serious issues in this area.

Summary

The development of citizenship education in the Netherlands has been most influenced by discussions on social integration and participation. This has led to aims for citizenship education being formulated in terms of making an active contribution and participation regardless of one's personal background. Through the statutory aims of citizenship since 2006, all schools are obligated to promote citizenship. Provision is complicated, however, as analysis shows discrepancies between general aims for citizenship and the core objectives for schools. This is further complicated by the freedom that schools are given to determine their own approach, which amplifies differences. The Inspectorate finds that the development of citizenship education is stalling, and for most schools, citizenship education appears to be low-key (Inspectie van het Onderwijs, 2011, 2012). Dutch students show civic knowledge scores close to the international average. However, scores on support for democratic values, equal rights and civic participation at school are all shown to be below the international average, leading to a problematic image of Dutch students' citizenship competences.

Citizenship education in Norway

Norwegian educational policy is largely decentralized through reforms over the past decades. Kindergartens, primary and lower secondary schools are operated by the municipalities, while county authorities have responsibility for upper secondary

education and training.[4] Over 60 percent of decisions are taken at the local level, compared to the OECD average of nearly 20 percent (OECD, 2013b). Schools have autonomy to interpret the national attainment targets set for the different subjects as well as in their choice of teaching materials (Mikkelsen & Fjeldstad, 2013).

Current educational policy is based on a number of key documents. The Core curriculum (Norwegian Directorate for Education and Training, 2006a) describes the fundamental values and views of humanity underlying education. It expands on six themes: moral outlook, creative abilities, work, general education, cooperation and natural environment. Based on these six themes, the document describes the student characteristics to

be fostered by education. The Quality framework (Norwegian Directorate for Education and Training, 2006c) summarizes and elaborates the central curriculum aims for schools and education as a whole. It explicates two school responsibilities related to citizenship education: social and cultural competence, and pupil participation. Subject syllabi set goals for what pupils should know after years 4, 7 and 10 and contain a distribution of teaching hours.

Citizenship education policy

A survey of human rights and democracy in the Norwegian curriculum concludes that all subjects can be said to cover democratic issues (Utdanningsdirektoratet, 2012). However, some subjects can be considered to be more closely aligned with these issues than others. Here we discuss the subjects of social studies, pupil council work, and religion, philosophies of life, and ethics.

Social studies ('samfunnsfag')

Throughout primary and lower secondary school, pupils are taught social studies through History, Geography and Sociology (and, since 2013, Exploring). "The purpose of the social studies subject is to help create understanding and belief in fundamental human rights, democratic values and equality, and to encourage the idea of active citizenship and democratic participation" (Norwegian Directorate for Education and Training, 2013). Though there is no formal decree that citizenship is primarily part of social studies, Mikkelsen & Fjeldstad (2013) propose that informally, social studies teachers are probably regarded as those best qualified to teach civic and citizenship education. Christophersen et al. (2003) found the revised social studies subject placed more emphasis on active participation by

---

[4] Approximately 50 percent of the kindergartens are privately owned. Municipalities must approve kindergartens and provide guidance to them.

students since the 1990s, but teachers appeared to maintain traditional ways of teaching. Pupils' political and democratic engagement appeared to be mostly influenced by factors outside school (e.g. media and parents), as teachers strongly emphasized the teaching of formal political institutions.

Pupil council work ('elevrådsarbeid')
Norwegian schools have a longstanding tradition of pupil councils. Pupil councils have been mandatory since the 1960s. Borhaug (2010) remarks that Norway seems to be the only country that elaborates on pupil councils in the national curricula – in addition to making them statutory by law. In lower secondary school, all pupil are taught 71 hours of 'pupil council work'. The subject focuses on developing pupils' ability to express their opinion and to function in various roles and groups, and developing pupils' understanding of democracy and participation in democratic processes. Pupil performance in the subject is not graded or tested. From 2014-2015, 'pupil council work' is longer a statutory subject but integrated with social studies. Additionally, pupils will be able attend the voluntary subject 'democracy in practice'.

Participating in a pupil council appears to have a positive effect on pupils' citizenship. Lauglo & Oia (2006) find very strong effects for participation in school council on all measures of civic engagement. While the presence of pupil councils is generally applauded, their focus and activities are not without criticism. Borhaug (2006, 2010) paints a picture of student councils doing valuable work and being positively regarded in schools by head teachers, teachers and students. However, he also finds their activities are mostly aimed at practical tasks originating from management and the contact teacher, and their practice does not adhere to democratic standards. In line with Borhaug, Solost (2011) finds that schools have untapped potential for student participation and activities relating to citizenship education. Notwithstanding these critical notes, *Norwegian students score high on civic participation at school,* indicating that the current approach is beneficial to students' participation.

Religion, philosophies of life, and ethics ('religion, livssyn og etikk')
Religion, philosophies of life, and ethics is taught in primary and lower secondary school. While religion takes a major portion of teaching time allocated to the subject, the values and virtues discussed through philosophies of life and through ethics can be considered relevant to citizenship education. These include competence aims such as "enable pupils to talk about ethnic, religious and ethical minorities in Norway, and reflect on the challenges of multicultural society"

(Norwegian Directorate for Education and Training, 2006b). While it is unclear how effective these aims are taught, the fact that they are not only part of the core curriculum but also an integral part of the curriculum can be considered to contribute to students' support for democratic values. Indeed, *Norwegian students score high on support for democratic values.*

Themes in citizenship education

Besides the discussions on citizenship education policy, a number of themes can be identified which can be considered to characterize citizenship education in Norway. We briefly discuss some of the most prevailing and typical of these.

Attention for citizenship

Pupil performance on the 2001 Civic Education Study shows generally favourable outcomes for Norwegian 14-year-olds (Mikkelsen et al., 2002). However, these positive results for students' citizenship competence at the start of the century do appear to also have a negative side. Stray (2009) concludes that relatively little attention is paid to citizenship education and democratization, as strengthening citizenship education is not an issue in Norwegian education discourse (see Mikkelsen & Fjeldstad, 2013). After the poor performance of Norwegian pupils on early PISA surveys and the relatively high scores of Norwegian pupils on the 2001 Civic Education Study, the focus shifted towards reading, mathematics, and science (Mikkelsen & Fjeldstad, 2013). This can also be seen to be reflected in the educational discourse surrounding the 2006 Knowledge Promotion Reform focusing on pupil performance, with less attention for the role of school as an arena for democratic citizenship. Given that explicit attention for citizenship competences can be considered particularly necessary for the development of citizenship knowledge, Norwegian students' citizenship knowledge could suffer. However, this is not reflected in the ICCS, as *Norwegian students score high on citizenship knowledge*.

Democratic participation

Considerable attention in the discussion of citizenship education in Norway goes to the type of political participation that is promoted. Somewhat paradoxically, Fjeldstad & Mikkelsen (2003) find Norwegian 14-year-olds score low on conventional types of participation but high on modern types, while educational policy and practice are aimed mostly at conventional citizenship, emphasizing normative issues and representative democracy (see Biseth, 2009; Samuelsson, 2013; Solhaug, 2003). Students' actual political involvement is thought to manifest

itself in more modern social expressions (Fjeldstad & Mikkelsen, 2003; Liden & Odegard, 2002; Rye & Rye, 2011). There appears to be little indication of students' dissatisfaction with politics and political issues, but also little engagement with these topics through conventional means. This trend is also found in the general population (Listhaug & Gronflaten, 2007), and is addressed through student elections (parallel to national elections) and a recent pilot allowing 16-year-olds to vote in one-quarter of municipalities in the 2011 local elections. Possibly reflective of this attention for democratic engagement, *Norwegian students score high on expected adult electoral participation.*

Equity and equality

Norway is renowned by the OECD for its higher-than-average scores in PISA and the low impact of socio-economic status on performance (OECD, 2004, 2013a, 2013b). Education policy restricts group setting based on abilities, sex or ethnicity except to respond to a defined

pedagogical need for a short time. While the OECD reports are generally optimistic, Norwegian researchers appear to remain critical of inequalities in the educational system and continue to stress its importance (see Opheim, 2004; Bakken & Elstad, 2012). The Norwegian Directorate for Education and Training (2008) maintains that to ensure equity in education for all, positive discrimination is required, not equal treatment. It appears then that within-country evaluations of equity in Norwegian education are critical and continue to emphasize the importance of increased attention, while in cross-country comparison Norway stands out positively. Interestingly, discussions on the integration of minorities and equal opportunities appear far less prominent in Norway than in other countries. The Directorate of Integration and Diversity (2009) concludes that public opinion towards immigrants in Norway is, in most cases, clearly on the liberal and tolerant side of the average. It appears then that Norway fosters a policy based on equitable beliefs, and support for equal opportunities appears evident. These values are also reflect in *Norwegian students scoring high on attitudes toward equal rights for all ethnic/racial groups.*

Summary

Students' citizenship competences in Norway are generally shown to be developing well. Education policy during the past decade has focused largely on cognitive outcomes, but mechanisms to promote students' citizenship can still be identified embedded in the curriculum. Discussions on young people's citizenship appear mostly to focus on the type of political engagement that is developed, though

arguably other aspects of citizenship have become so evident they are hardly problematized. Norwegian students score high on all aspects of citizenship competences included here.

Citizenship education in Scotland

The development of citizenship education in Scotland is often considered in comparison to developments in England. However, although citizenship education in England has definitely been of influence, the development of 'education for citizenship' in Scotland has been distinctively – and continues to be increasingly – different from the English case (Kerr et al., 2008). Notably, while attention for citizenship education in England has focused on compensation for the 'democratic deficit', non-cognitive outcomes such as citizenship have from the outset been conceived as crucial components of Scotland's 5-14 program in the early 1990s (Carr, 2003).

A general discussion of the state of education for citizenship in Scotland is complicated by the fact that there has yet to be any major research in Scotland on the implementation of education for citizenship across the country (Munn & Arnott, 2009). While national curricula are set centrally, most decisions are taken at the local or school level, comparable to the OECD average (OECD, 2012).

Citizenship education policy

Early this century, the National Advisory Group 'Education for citizenship in Scotland' published two papers on the implementation of education for citizenship in the Scottish curriculum (LTS, 2000, 2002). The papers provided the official framework for education for citizenship. They proposed a view of young people to be regarded as citizens of today rather than citizens in waiting and advocated an active approach that enables young people to act and participate in various communities. Several years after the framework for citizenship, the Curriculum for Excellence (CfE) was introduced. The new curriculum is generally considered to be in line with the ambitions of the 2000 education act and the 2002 education for citizenship framework (HMIE, 2006a; Munn & Arnott, 2009).

Evaluation by the Inspectorate in 2006 showed that schools are gradually improving their citizenship practices (HMIE, 2006a, 2006b). Schools are generally found to have increased their emphasis on citizenship education. However, practice is uneven across schools, and some note that more effort is needed to fulfil the aspirations that were set out (Kisby & Sloam, 2012).

Themes in citizenship education

A number of themes can be identified which can be considered to characterize education for citizenship in Scotland. The following section describes some of the prominent themes in education for citizenship policy.

Participation

Earlier approaches to education for citizenship emphasized participation as a social aspect of citizenship, with less attention for the political and democratic dimension (see Biesta, 2009; Munn & Arnott 2009). This claimed lack of attention for politics is remarkable considering youth's reported political disengagement (Maitles, 2000; Munn et al., 2004). Recently, however, political participation appears to have gained more attention (see Education Scotland, 2013a). Given the Scottish government's intention to allow 16 and 17-year-olds to vote in the 2014 referendum on Scottish independence, it is considered essential for young people to develop their political literacy and engagement (Education Scotland, 2013b). A recent survey finds that young people in Scotland do not appear less interested in politics than the overall population and are largely willing to participate in the referendum (Eichhorn, 2013). These findings lead one to expect students to have a general affinity with political matters but little active engagement, thus *Scottish students are expected to have average expected adult electoral participation.*

Pupils are expected to develop a positive attitude by being engaged with school life and participating in decision-making in the school community (LTS, 2007; Maitles & Deuchar, 2007). Pupil participation in school is, moreover, believed to foster effective citizens and active community involvement in later life (Deuchar, 2003; HMIE, 2006a). While the attention for school ethos and pupil participation has not been without problems, there appears to be broad support for their importance as a way of promoting students' citizenship competences. Cowan & McMurtry (2009) found that by 2004 about a quarter of schools had established a progressive approach to education for citizenship, while at the other end of the spectrum a quarter of schools appeared to take a minimalist approach. By 2008, Cross et al. (2009) found that the vast majority of schools had developed plans to encourage pupil participation and emphasized opportunities for pupils to be involved in decision-making over other forms of participation. Assuming this trend has continued, *Scottish students are expected to score high on civic participation at school*.

School councils

Earlier this century, the most common approach toward a participative school ethos was the establishment of a school council (Deuchar, 2003; HMIE, 2006a). In 2002, 83 percent of primary schools and 96 percent of secondary schools had a School Board (Scottish Executive,

2002). School Boards were abolished and replaced by school councils in 2006 (Scottish Executive, 2006). Most schools have established a school council to allow pupils to participate in decision-making (Ross et al., 2007). These are generally aimed at improving practice rather than challenging existing systems (Munn & Arnott, 2009). However, while senior students in both primary and secondary schools are actively engaged, this is argued to be much less the case for younger students in each school type (Deuchar, 2009; LTS, 2007). Only in half of the schools are all years represented in the school council (Cross et al., 2009). There is mixed evidence of the effects of school councils: while HMIE (2006a) finds that schools are increasingly realizing positive effects, other research observes that practice is predominantly weak and few students have a say in what is taught and how they want to learn (Maitles & Deuchar, 2007; Mills, 2004; Potter, 2006). Considering that the attention for school councils seems less prominent in contemporary documentation, and given the different findings considering effectiveness, *Scottish students' civic participation at school is expected to vary across schools.*

Modern Studies

Modern Studies is offered as a social subject – beside History and Geography – and is considered to pose a particular distinctiveness to the teaching of education for citizenship in Scotland (Munn & Arnott, 2009). Overall, learning and teaching in social studies is found to be strong, effective and improving, with ample examples of good practice (Education Scotland, 2013c). An inventory of the possible contribution of the social subjects to citizenship education shows that all three subjects offer significant opportunities for the development of knowledge and understanding and skills relevant to citizenship (Munn et al., 2004a). While the contribution of Geography and History varies greatly, Modern Studies offers a whole-school citizenship resource, particularly in political literacy (Education Scotland, 2013c; HMIE, 2007). Some even go so far as to say the existence of Modern Studies has meant an evolution in approaches to citizenship education (Kerr et al., 2008). However, since Modern Studies is not compulsory, only one in three students on average is likely to receive formal citizenship education (Andrews & Mycock, 2007). Consequently, students who attend Modern Studies

can be expected to score favorably on citizenship knowledge, but those who don't could very well be at a disadvantage. Again, since there is a lack of insight into students' citizenship knowledge performance, these conclusions remain tentative. The expectation here is that *Scottish students score average on citizenship knowledge.*

Equity and equality
Equity and equality have been considered a fundamental characteristic of Scottish education. While the workings of these features have been questioned, they are still thought to influence education and education policy (Freeman, 2009; Munn & Arnott, 2009; Priestley & Humes, 2010). Whether or not Scottish education achieves equitable results is a matter of debate (see OECD, 2007), but the focus on these fundamental values indicates that they remains a key ideology of the Scottish education system (McCrone, 2003). General support for equality can therefore be expected to somehow permeate education and consequently affect students' value development. Immigration in Scotland appears low-key, with close to 95 percent of students in primary and secondary schools classified as 'White-UK' (Scottish Government, 2013). The Scottish Social Attitudes survey shows that most Scots are supportive of ethnic diversity, but there is a substantial minority who hold discriminatory views (Ormston et al., 2011). These views are found to have slightly changed in a negative direction over the past years, though the authors remain optimistic that views are more tolerant with younger generations and increased education attainment. Considering that equality appears to be a fundamental value in education, that immigration is low and that young people score relatively high on ethnic tolerance, *Scottish students are expected to score high on attitudes toward equal rights for all ethnic/racial groups.*

Summary
Education for citizenship in Scotland is promoted in a cross-curricular manner as part of the Curriculum for Excellence and social studies. The aims for development of citizenship are one of five central priorities of education. Education for citizenship in Scotland can be characterized by attention for student participation, the position of school councils, the 'Modern Studies' subject, and the value of equity and equality. Based on the account given above, Scottish students are expected to score average on expected electoral participation, high on civic participation at school, average on citizenship knowledge, and high on support for equal rights. Furthermore, students' participation at school is expected to vary

significantly between schools. Unfortunately, there is currently no data available to further examine these expectations.

Citizenship education in Sweden

Much of how citizenship and citizenship education in Sweden are thought of seems to have its roots in the social democratic model of Swedish society since the end of the Second World War. The developments that have taken place since then have shaped the way development of citizenship among young people is promoted. Though Sweden has traditionally placed much
emphasis on equality and uniformity, the present reform cycle shows these are gradually replaced by diversity and liberty which has increased school segregation and differences between pupils (Arnesen & Lundahl, 2006; Lundahl, 2002; Skolverket, 2013). The National Agency of Education recognizes four themes in recent development tendencies: segregation, decentralization, streaming and individualization (Skolverket, 2009b). Evaluations indicate growing differences in grades and achievements related to gender, social and ethnic background (Lundahl, 2005), and large variations between municipalities (Skolverket, 2009b). Education is now much more recognized as a sorting mechanism than before (Lindblad et al., 2002), and school-level effects have been reported to have intensified (Skolverket, 2006). In terms of levels at which educational decisions are taken, Sweden is very close to Scotland and comparable to the OECD average (OECD, 2012).

Citizenship education policy

The changes in Swedish education over the past decades have had a large impact on the way citizenship among young people is viewed. Stolz (2011) argues that Swedish education policies no longer emphasize equality aimed at individual children, and while citizenship was emphasized in the 1968 and 1991 upper secondary education reforms, democracy or citizenship are left unmentioned in the 2009 reforms (Lundahl et al., 2010). Bernitz (2012) concludes that citizenship is currently not a big political issue and that a weakening of the concept has occurred. The current national school curriculum stresses citizenship development in terms of norms and values. It directs that "schools should actively and consciously influence and stimulate pupils into embracing the common values of our society, and their expression in practical daily action" (Skolverket, 2011a). A number of goals and guidelines are set for schools, based on fundamental values and tasks of the school.
The overarching goals and guidelines do not, however, directly correspond to the perceived practice in the classroom. Aldenmyr et al. (2012) typify Swedish

education policy as neo-liberal for the stress that is put on developing autonomy through choice. At the level of teachers, these authors find this to be overshadowed by the opinions of what teachers believe the students should become. This leads them to conclude that the envisioned citizen is one who is active, competent and self-made in a certain fixed and pre-defined way. The ideal citizen is, then, one who shows exemplary behavior without addressing fundamental questions of equity and authority (see Arensmeier, 2010; Skolverket, 2004).

Themes in citizenship education
Besides the discussions on citizenship education policy, a number of themes can be identified which can be considered to characterize citizenship education in Sweden.

Civics ("samhällskunskap")
Citizenship is offered as a separate subject of Civics and a cross-curricular theme through whole school promotion of the fundamental goals and objectives specified in the curriculum. Civics is specified in terms of subject aims as well as core content and knowledge requirements. The 2000 and 2008 syllabus for Civics formulated most aims for pupils in terms of developing knowledge (Skolverket, 2009a). The 2011 curriculum still features a knowledge component, but added a set of ability aims. Pupils are to be equipped with the tools necessary for dealing with the attitudes and values of citizenship education, but how teachers should become equipped in delivering these is not specified (Sandstrom Kjellin & Stier, 2008). However, this does not appear to be of concern, as teachers of social studies are reported by most pupils (88 to 90 percent) to teach well (Skolverket, 2004). As the earlier syllabi for Civics place the most emphasis on the development of knowledge, and as teachers highly prioritize knowledge of civic rights and obligations (Ljunggren & Ost 2010), provision in this domain can be expected to be fairly effective. Indeed, *Swedish students score high on citizenship knowledge.*

Democratic values
Citizenship education in Sweden has been noted from an international comparative viewpoint to place much emphasis on the education of values. The emphasis on values is reflected in the fact that citizenship education is often referred to as 'values-based education' in policy documents (Mikkelsen, 2004; Council of Europe, 2005). Though this focus on values in policy documents has been acknowledged, it is also noted that the course literature focuses on democracy as a decision-making process (Bernmark-Ottosson, 2005). This inconsistency between policy and practice has also been noted from a teacher's point of view. In 2000,

only approximately half of the teachers felt that school is successful in communicating basic values and traditions (Skolverket, 2001). Staff and students have a high awareness of the values system, but nonetheless evaluations indicate clear deficiencies in basic values application in everyday school activities (Skolverket, 2004). Lack of explicitness of democratic values and principles is also found in schoolbooks (Arensmeier, 2010). Civic education is a normative subject that could result in normative dilemmas when wanting to discuss different political and normative values (Ljunggren & Ost, 2010). Teachers have been noted to consider themselves unable to use appropriate strategies to do this (Skolverket, 2001).

The cause for this inconsistency between policy, practice and the perceived high levels of awareness of values could be envisioned to lie outside the school. The high awareness of values can be explained when the school is envisaged not as a place where learning about society takes place, but as a public sphere that is influenced by the society in which pupils participate (Amnå et al., 2010). The high level of value awareness outside the school permeates into the classroom. Swedish students have been found to consistently and strongly support democratic principles (Eriksson, 2006). Somewhat surprisingly then, *Swedish students score average on support for democratic values*. This could indicate that schools and teachers are indeed experiencing difficulty in promoting these values.

Equality
Arensmeier (2010) finds that youth's support for equality is obvious (cf. Eriksson, 2006), but not explicitly voiced. This could be because these values are taken for granted by the students. Though the concept is not explicitly voiced, it can be regarded as being an integral part of Swedish society as similarly discussed for democratic values. The social democratic model introduced after the Second World War particularly promoted equality, and it is thought to still be prevalent in present-day society and thus reflected in schools (Kerr et al., 2010). ICCS results confirm this picture showing *high support for equal rights for all ethnic groups among Swedish students*.

Democracy in schools
In 2002, students' right to participate in school planning, even at young ages, was enhanced (SOU, 2002). This is reiterated in the 2011 curriculum, which stated that teaching should not only impart knowledge about fundamental democratic values, democratic work forms should also be applied in practice and prepare pupils for

active participation in society (Skolverket, 2011a). Studies on participation in schools, however, show contradictory findings.

At the school level, students do not regard school to be a democratic arena but instead talk about education for democracy in later life, after their school years (Arensmeier, 2010; Eriksson, 2006). Opportunities for pupils to exercise influence and to have an impact on schools are relatively limited (Skolverket, 2009b). Democracy in schools was found to conflict with schooling for democracy, where student influence had a negative effect on the fostering of values and equality of citizenship education (Almgren, 2006). Democratic opportunities were found to differ among pupils, which increases inequality.

At the classroom level, the picture is somewhat more optimistic. The majority of upper secondary students consider the classroom climate democratic (Skolverket, 2003). Teachers are generally sympathetic to student participation and school democracy (Skolverket, 2011c). Ljunggren and Ost (2010) even conclude that education consists of a communication climate mainly characterized by free speech and open discussion. These combined findings present a picture in which whole school democracy appears problematic, but teachers appear to be supportive of fostering a democratic climate. While the selected ICCS scale considers the school level, *Swedish students score average on civic participation at school.* This could indicate that the dichotomy illustrated above between classroom and school level is not a strict separation, and that participation occurs at both levels.

Political participation

Based on interviews with ten focus groups of young people, Arensmeier (2010) finds that young people support the idea of democracy but that their feelings towards politicians are mainly negative. This finding also comes forward in other discussions, as Eriksson (2006) reports politicians are frequently discussed in her interviews, and the tone of such comments tends to be critical. Bernmark-Ottosson (2005) concludes that young people in Sweden are generally interested in politics. The image is complex, however, as the percentage of young people who vote in public elections is decreasing – albeit still higher than in most other countries – as is the readiness to engage in traditional party politics, while interest in single-issue movements is increasing. This gap between young people's single-issue interest and traditional politics means that the youth are increasingly distanced from democratic participation and political activism. This gap does not appear to be bridged by citizenship education, as Amnå et al. (2007) consider Swedish politics to be marked by an intertwined social and political concept of citizenship, but attention must now focus on the political side of citizenship. Students'

disengagement with traditional politics is confirmed, as *Swedish students score low on expected adult electoral participation.*

Summary
Sweden is noted for its promotion of a highly social political program after the end of the Second World War, which has been shown to largely align with the values and ideals of active citizenship. Much has happened since, however, as the focus of education shifted to the individual in the 1970s and liberal influences increased throughout political and educational policy. Though compared to fifty years ago, Swedish policy is now much more liberal, it is still noted from a European perspective to have a strong social character. As these values can by now be expected to have perpetrated daily life, they are expected to have positively influenced citizenship development among young people; both inside and outside of schools.

Overall, Sweden has traditionally been one of the high-scoring countries on citizenship in Europe. Contemporary changes have shown to somewhat negatively influence outcome measures, but Swedish students still perform well in international comparison. Swedish students score high on citizenship knowledge, average on support for democratic values, high in support for equal rights, average on civic participation at school, and low on expected electoral participation.

Students' citizenship competences in cross-national comparison

This section considers in some more detail the ICCS results for students from the Netherlands, Norway and Sweden.[5] Table 1 presents the mean scores and standard deviations on the selected scales for each country. The scores are normalized internationally to have a mean of 500 and standard deviation of 100 for citizenship knowledge, and a mean of 50 and a standard deviation of 10 for all other scales.

Possible causes and implications of students' performance are discussed in the preceding country sections. The results generally show clear differences between students' citizenship competences in the selected countries. The overview shows that Dutch students score significantly below the international mean on all scales except citizenship knowledge. Norwegian students on the other hand score significantly above the international mean on all selected scales. The image is more mixed for Swedish students, who score above the mean on knowledge and support for equal rights, but below the mean on electoral participation. The differences are

---

[5] As mentioned before, Scotland did not partake in the ICCS 2009 study.

generally small but statistically significant. For a more detailed analysis of students' citizenship competences based on ICCS results in these countries, we refer to the respective country reports for the Netherlands (Maslowski et al., 2012), Norway (Mikkelsen et al., 2011) and Sweden (Skolverket, 2011b).

Table 1. Average student citizenship competences per country (ICCS 2009)

	Netherlands			Norway			Sweden		
	Mean	sd.	ISC	Mean	sd.	ISC	Mean	sd.	ISC
Citizenship knowledge	493.61	91.10	.59	514.87 △	95.97	.12	531.02 △	97.64	.15
Democratic values	46.25 ▼	9.28	.08	50.63 △	10.31	.02	50.46	10.81	.06
Equal rights	47.07 ▽	10.01	.08	51.63 △	10.57	.03	52.00 △	11.17	.07
Electoral participation	46.56 ▼	9.69	.13	51.83 △	10.51	.05	49.01	9.59	.07
Participation at school	42.70 ▼	10.65	.18	54.02 ▲	9.75	.04	50.43	9.60	.09

▲ More than 3 score points above ICCS average (excluding knowledge)
△ Significantly above ICCS average
▽ Significantly below ICCS average
▼ More than 3 score points below ICCS average (excluding knowledge)
ISC: Intra-school correlation (.05 small; .10 medium; .15 large (cf. Hox, 2002))

School effects

While the ICCS assesses individual students' citizenship competences, multilevel analysis allows us to investigate to what extent these results can be ascribed to differences between schools (see Hox, 2002). The intraschool correlation (ISC) in Table 1 can be interpreted as the proportion of students' citizenship competences that can be considered common to students attending the same school.[6] In other words, the ISC indicates whether it makes a difference if students attend one school or another. While the question of whether these differences are the result of school effort or pre-existing difference in student population remains, the ISC does provide some indication of school effectiveness.

A number of interesting results emerge from the ISCs presented in Table 1. First, differences between schools are the largest in the Netherlands and the smallest in Norway. This corresponds to large differences in school autonomy between the selected countries (see OECD, 2012), and is likely (also) influenced by the differentiated structure of Dutch education. Second, all countries show a

---

[6] ICCS sampled one class per school, so the ISC represents both class- and school-level variation. In the Netherlands, some schools sampled two classes, which could affect the estimates in Table 1.

medium to large effect size on citizenship knowledge. This can be considered an indication that schools have a considerable effect on students' citizenship knowledge development, and that effectiveness differs between schools (the effect size for the Netherlands is very large, likely magnified by differences in citizenship competences between educational tracks; see ten Dam & Volman, 2003). The effect sizes for citizenship knowledge are comparable to the analysis by Schulz et al. (2010), who find ISCs between .6 and .52 across countries, with an average of .28.[7] In comparison to these figures, the ISC for citizenship knowledge can be considered small in Norway and Sweden and large in the Netherlands. Third, differences in students' participation at school show medium to large differences between schools in Sweden and the Netherlands respectively, which indicates schools differ in the extent to which they (effectively) provide for this. Finally, ISCs are small for democratic values and equal rights, which could mean two things: either schools' approach to development of these values are very similar within countries, or students' development of these values occurs mostly independent from school. Which explanation is chosen also affects the approach to the assessment of school and students. If schools are believed to have little effect on the development of values but a large effect on knowledge, this has implications for the type of assessment that can be deemed suitable.

Measuring and assessing citizenship competences
The type of assessment employed by surveys such as the ICCS is not without criticism. The approach taken can be considered similar to other international studies such as PISA. Some would argue such an approach to measuring citizenship competences fails to capture the complex, contextual nature of the development of citizenship. Given that the preceding analysis showed that effectiveness in citizenship education differs between schools, this brings up the question what approach best allows us to capture and support both student and school development (see Dijkstra & de la Motte, 2014).

In Scotland the use of outcome measures to judge school and pupil performance appears to be considered undesirable (HMIE, 2006b; Munn et al., 2004b). While assessment outcomes are shared with schools to compare performance, these are not published nationally (Scottish Executive, 2009). Instead, the Inspectorate takes an enabling approach by promoting a culture of self-evaluation whereby schools improve their own performance through provided audit

---

[7] The analysis by Schulz et al. (2010) only featured citizenship knowledge and excluded the Netherlands, as the country did not meet the required sample size.

materials. This contrasts with the approach taken for example in the Netherlands, where insights into the quality of schools' citizenship education are sought after (Ministerie van Onderwijs, Cultuur & Wetenschap, 2013; Onderwijsraad, 2012). The approach taken to assessing schools and students has consequences for the type of instruments that can be considered most suitable (Dijkstra & de la Motte, 2014). It seems evident that an instrument that tests students' knowledge compared to a set standard is less suitable to provide schools with practical directions for school improvement than feedback provided by the Inspectorate (see Dijkstra & de la Motte, 2014).

Discussion
To support the analyses conducted in this chapter, five of a total 25 scales from the ICCS were selected. The analysis rests on the assumption that the selected scales present an accurate measure and are representative of the notions discussed. Results might have been different had different scales been selected, but the results showed a coherent picture overall. By combining a synthesis of citizenship education in each of these countries with student performance data, this chapter aimed to further add to the growing knowledge base on (the development of) citizenship education. While the context of this chapter did not allow for an exploration of all relevant complexities, it is believed that the images portrayed provide an accurate general account.

CHAPTER 3: ASSESSING CITIZENSHIP COMPETENCES USING A RUBRIC-BASED APPROACH[1]

Abstract

Assessing young people's citizenship competences is a challenging endeavour. Large-scale surveys play an important role, and mainly focus on assessment *of* learning. Instruments supporting assessment *for* learning are scarce. Various studies have suggested rubrics are a possible assessment instrument that could facilitate both assessment *of-* and *for* learning. In this paper we evaluate the development of a rubric-based approach to measuring citizenship competences. The central issue is: To what extent can rubrics support assessment of students' citizenship competences? Using data obtained from 716 students in the fourth year of general secondary education (grade 10) and the first year of vocational tertiary education (grade 11), we developed and tested several steps towards a theoretical and operational framework of rubrics relevant to students' citizenship in everyday life. Further research into using rubrics for self-assessment of citizenship attitudes might be particularly worthwhile, considering the potential this could offer to support future learning.

Introduction

Assessing young people's citizenship competences has proven a challenging endeavor (cf. Richardson, 2010). For young people to take an active part in various communities as well as in general society, they need to develop the necessary knowledge, attitudes and skills – that is, citizenship competences – to do so. Assessing students' development in this area can help to support their learning process *and* to evaluate the effectiveness of educational programs. However, the concept of citizenship can be considered fluid, in the sense that the discourses regarding the meaning of citizenship reflect and change with the context in which citizenship is considered (cf. Knight-Abowitz & Harnish, 2006). Citizenship refers to a plethora of contexts and thus manifests itself in various ways, and the competences necessary may vary accordingly. Moreover, it plays a role in diverse social domains (in relationships between people and groups, and in the relationship between people and groups and the government) and comprises a normative

---

[1] Based on: Daas, R., Dijkstra, A. B., Karsten, S., & Ten Dam, G. (submitted). Assessing citizenship competences using a rubric-based approach.

component. The challenge for any assessment of citizenship competences is to do justice to these varying contexts and ways in which citizenship can be considered.

The approach taken by most large-scale surveys is to adopt a broad scope on citizenship competences. Examples of these survey approaches are the Civic Education Study (CIVED), the International Civic and Citizenship Study (ICCS) and the Citizenship Education Longitudinal Study (CELS; Keating, Kerr, Benton, Mundy, & Lopes, 2010; Schulz, Ainley, Fraillon, Losito, & Agrusti, 2008; Schulz, Ainley, Fraillon, Losito, Agrusti, & Friedman, 2017; Torney-Purta, Lehmann, Oswald, & Schulz, 2001). Such approaches assess citizenship competences on a wide range of items. The items aggregate to citizenship knowledge scales and a set of attitudinal scales, to do justice to the range of content domains as well as varying social and political settings in which citizenship can take shape.[2] The CIVED and ICCS studies are cross-national, allowing comparisons between students from up to 38 different counties. The CELS is one of few large-scale longitudinal studies of the development of citizenship competences, and was conducted in England. These studies have played an important role in furthering our understanding of students' citizenship competences. However, they also reflect the complexity of the concept. Hoskins, Villalba, and Saisana (2012) attempted to create a single composite indicator for citizenship competences based on ICCS data, but find 17 country-specific frameworks for citizenship competences based on the data from 24 European countries.

Although they are based on different conceptual frameworks, the CIVED, ICCS and CELS have a lot in common. They define citizenship competences on the basis of a sense of 'agency' on the students' part, and consider students to be part of multiple civic communities with which they interact, both now and in the future. The nature of these interactions varies according to the community, since students' opportunities for involvement differ according to their background, resources and experiences. The studies also employ similar methods. They (predominantly) use a combination of a multiple choice test and questionnaire, combining various items to form meaningful scales. Students' scores on these scales can be used to compare their competences.

Multiple-choice tests and questionnaires can be considered well suited to assess citizenship competences efficiently, and to result in a picture of the students'

---

[2] The CIVED and ICCS frameworks include 10 attitudinal outcome scales for citizenship competences and 2 subscales for civic knowledge. Although CELS focuses on 8 student level outcomes of citizenship education, a range of scales is employed throughout the study.

level of proficiency. Tests and questionnaires are particularly useful to compare (groups of) students in these respects. Large-scale studies employing these instruments have shown to be valuable for a range of purposes, enabling, for example, comparisons between students or between schools and identifying school characteristics that coincide with high levels of citizenship competences. However, the context with which students are provided is often limited, as are opportunities for them to say more about topics they consider more or less relevant. These instruments are therefore arguably less suited to reflect the contexts in which students' citizenship competences take shape, nor do they provide an understanding of the students' reasoning or arguments for (dis)agreeing with a certain statement (see Chapter 5).

These instruments can be considered to primarily facilitate assessment *of* learning. Assessment of learning aims to evaluate the students' level of competence or the extent to which the intended learning outcomes have been achieved. The results can be used to describe the students' current knowledge or attitudes or to evaluate their achievement or progress, but also serve more broadly for accountability purposes in education (e.g. pass/fail grades for students, comparing schools or education systems in benchmark-oriented approaches or school accountability; cf. Dijkstra & de la Motte, 2014). By contrast, assessment *for* learning focuses on informing students' future learning and supporting their development of knowledge or skills (Black & William, 1998; Broadfoot et al., 2002). Although instruments aimed at assessment for learning citizenship competences could contribute towards promoting citizenship among students, only a handful are currently available (see Chapter 5).

Given that large-scale tests and questionnaires are prevalent tools for assessing citizenship competences, it would be worthwhile to consider alternative approaches aimed at supporting the learning process. A portfolio-type approach appears promising to complement the instruments currently available (see Chapter 5; Kerr, Keating, & Ireland, 2009). Because portfolios provide a more detailed understanding of student citizenship learning and their considerations, they are considered better suited to provide input to promote learning, that is, assessment for learning. However, because of this inherent link to student learning, portfolio assessment can be considered both an assessment instrument and a tool for learning (cf. Dierick & Dochy, 2001; Jerome, 2008). In this study, we will explore these possibilities by developing a new assessment instrument for citizenship competences and focus on the development of an instrument that could facilitate a portfolio approach: rubrics. Rubrics potentially provide a basis for supporting a portfolio approach by helping to define and distinguish different levels of

performance (Johnson, Mims-Cox, & Doyle-Nichols, 2010). We aim to develop a rubric instrument for the assessment of citizenship competences, focusing on the viability of assessing citizenship competences through rubrics, which would add to the available instruments by providing a novel approach to the assessment of citizenship competences. The theoretical framework and methods sections will outline how we developed a set of rubrics used to assess students' citizenship competences. We aim to answer the question: *To what extent can rubrics support the assessment of students' citizenship competences?*

Theoretical framework

Citizenship competences

Citizenship competences can be defined in different ways. However, these different conceptualizations share common grounds on which to build. Fundamentally, citizenship competences involve more than acquiring knowledge of e.g. civic institutions or the separation of powers. They also include development of skills to take an active part in societal life and developing an attitude towards others. In other words, citizenship competences refer to (young) people's ability to live together in different roles, in relation to other individuals and groups, and as citizens interacting with government (cf. Rychen & Salganik, 2003). This study builds on the conceptualization formulated by Ten Dam et al. (2011), which focuses on the assessment of citizenship competences as relevant to students' daily lives. To assess students' capabilities to take an active part in society, these authors consider young people's citizenship as manifested in four exemplary social tasks: acting democratically, acting socially responsible, dealing with conflicts and dealing with differences. Social tasks refer to tasks that anyone – including young people – involved in society will have to fulfil in order to function in a group or in society as a whole (Ten Dam & Volman, 2007). Better performance of these tasks depends on the students' relevant knowledge, attitude and skills. Citizenship competences are thus measured by assessing students' knowledge, attitudes and skills pertaining to each social task.

The emphasis on social tasks also underlines the relevance of the context in which citizenship competences are used. Because our study is conducted in the Netherlands, this concerns the societal context relevant to Dutch students. However, the themes playing a role here – such as diversity, democracy, equality and inequality and social trust – are not that different from those in many other democratic societies (Eurydice, 2017). As in many other countries, the debate on citizenship education in the Netherlands is closely linked to notions of social

cohesion (see Chapter 2; Bron, 2006). Although students will generally not have reached voting age, they are commonly regarded as members of society within the context of the citizenship education debate, since they interact with their socio-political environment and other people around them.

As in many other countries, promoting citizenship competences is a statutory obligation for Dutch schools, and the objectives are stipulated in legislation. Moreover, Civics has been a school subject in upper secondary education since 1968, and includes topics such as 'parliamentary democracy' and 'the diverse society'. Civics is assessed by central examination. For citizenship education schools are free to choose how to assess whether students have met the requirements stipulated in the national learning aims. However, the Dutch Inspectorate of Education finds that schools lack insight into students' development in relation to these learning aims, and learning outcomes are not systematically evaluated (Inspectie van het Onderwijs, 2016).

Rubrics

Rubrics are a relatively new assessment instrument, and present a novel approach to the assessment of citizenship competences in particular. Rubrics are essentially a matrix consisting of two dimensions: the assessment criteria and the levels of proficiency. Each of the cells in this matrix specifies a description of expected performance. Rubrics are used to clarify learning goals, communicate those goals to students, guide student feedback to further their progress and assess learning outcomes (Andrade, 2005). Rubrics are considered to promote learning by making expectations and criteria explicit and by facilitating feedback and self-assessment, although the available evidence showing these benefits is scarce (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). The level of detail in the performance descriptions can range from broad descriptions in holistic rubrics (e.g. 'The student shows a willingness to help others'), to detailed scoring characteristics in analytic rubrics (e.g. 'Over the past week the student has helped at least two other students in class').

Holistic rubrics consist of only a few dimensions on which a task is evaluated, with broad descriptions for each performance level. These rubrics focus on overall development and allow for deviating student responses, suggesting there is no single correct answer (Mertler, 2001). Holistic rubrics aim to leave the construct that is being assessed intact, thereby retaining its meaning. At the same time, holistic rubrics leave room for interpretation, and require assessor training to reach agreement of interpretation (Van den Bos, Burghout, & Joosten-ten Brink, 2014).

Analytic rubrics divide a task into singular criteria which are described as clearly as possible. Each criterion is evaluated separately. These rubrics may consist of a list of criteria, each of which is systematically divided over several levels of performance (typically three to five). For analytical rubrics in particular, consistency of performance criterion descriptions is considered pivotal for reliable assessment (Tierney & Simon, 2004). These rubrics are more likely to produce agreement between assessors. Analytic rubrics also convey to students that a task is composed of a multitude of components. Although analytical rubrics are more time consuming than holistic rubrics, they are considered more effective to identify students' strengths and weaknesses (Jonsson & Svingby, 2007).

Many variations to these two archetypes are possible, and they present two ends to a continuum rather than a dichotomy. The choice of rubrics depends on the purpose for which they are used. Rubrics can support assessment *for* learning by specifying (different levels of) expected performance. Both students and teachers gain an understanding of possible avenues for further learning, since such assessment provides both transparency of proficiency levels and feedback on current ability (or 'feed forward', when the rubric is seen as the starting point for a learning process; cf. Hattie & Timperley, 2007). Using rubrics to facilitate assessment for learning commonly involves student self-assessment (Panadero & Jonsson, 2013).[3] To the best of our knowledge no research has been conducted into the application of rubrics to assess citizenship competences. A key question is therefore if students are able to use rubrics to assess their own citizenship competences.

When rubrics are used for (high-stakes) assessment *of* learning, reliability of scoring is particularly important, as is clarity of the assessment criteria and descriptions. Assessment of learning using rubrics generally involves assessment by others, for example when rubrics are used as a tool for scoring student work. Rubrics are considered a suitable tool for assessment of 'complex performances' (Jonsson & Svingby, 2007), to which we believe citizenship competences can be counted. We therefore evaluate if rubrics can be used to assess students' citizenship competences based on their work.

---

[3] There is ongoing debate on the use and definitions of the terms self-assessment, self-evaluation, or self-reflection (cf. Andrade & Du, 2007). In line with other studies on rubrics we use self-assessment to refer to students' evaluation of their own learning, achievement, or competences (cf. Panadero & Romero, 2014).

In this study, we aim to provide a systematic evaluation of employing rubrics to assess citizenship competences. Because the instruments available to date consist of a combination of tests and questionnaires these provide a benchmark to which we can compare our instrument. We therefore also investigate whether using rubrics leads to comparable results to using a test and questionnaire. To answer the central question posed in the previous section – *To what extent can rubrics support the assessment of students' citizenship competences?* – we will answer the following sub-questions:

- To what extent do rubrics support self-assessment of citizenship competences?
- To what extent do rubrics support assessment of citizenship competences based on students' work?
- To what extent does assessment of citizenship competences using a rubric produce similar outcomes to using a test and questionnaire instrument?

Methods

In order to evaluate the feasibility of assessing citizenship competences using rubrics, one of the main undertakings of this study was the development of a rubric instrument. We first describe the development of the instrument and the selection of participants. We then turn to the approaches taken to address each of the research questions.

Development of citizenship rubrics

For the purpose of this study, we designed a set of rubrics aimed at assessing citizenship competences. Based on the aforementioned conceptual framework developed by Ten Dam et al. (2011), we developed four frameworks of rubrics simultaneously for four exemplary social tasks relevant to students' citizenship in everyday life: acting democratically, acting socially responsible, dealing with conflicts and dealing with differences. As we discovered that dealing with conflicts is not an explicit part of the curriculum, we decided not to develop this framework further. Each framework described the social task in terms of knowledge, attitude and skills (dimensions) at four levels: A through D. For each cell in each of the frameworks, we developed a general description, and several examples of typical knowledge, attitudes or skills. The contents are based both on the theory of increasing levels of competence (cf. Bloom, 1956) and existing measures of

student citizenship competences (cf. Keating et al., 2010; Schulz et al., 2008; Ten Dam et al., 2011; Torney-Purta et al., 2001).

The frameworks were developed over the course of two years, during which researchers, teachers and students were consulted for input and feedback. The first versions of the frameworks were discussed with eight civics teachers in Dutch secondary schools and with six panels of students in the second through fourth years of secondary education (grades 8 through 10). The discussion with both teachers and students showed that younger students (aged 13 or 14 year) had considerable difficulty understanding the purpose of the rubrics. They appeared to consider the framework as a checklist, and had difficulty relating its content to their own experiences. This would mean rubrics would have little added value over a questionnaire. We decided to focus on students of 15 years and older and made several formatting changes to discourage students from treating the framework as a checklist. An added benefit of focusing on students of 15 years and over is that students in fourth-year general secondary education are taught 'civics' and students in first-year vocational tertiary education are taught 'career and citizenship'.

We tested the revised version of the frameworks in a pilot involving nine civics or career and citizenship teachers teaching fourth-year general secondary education (grade 10) or first-year tertiary vocational education (grade 11). Students from eleven classes assessed themselves using one of the frameworks. Students from nine classes completed a reflection assignment inspired by one of the frameworks; their reports were assessed using the scales for knowledge and attitude. This pilot provided valuable input, based on which we changed some of the phrasing to clarify the structure and differences between levels, which is considered one of the most challenging aspects of designing rubrics (Reddy & Andrade, 2010). It also provided input to re-evaluate whether the examples in the framework reflected the corresponding level of competence. Because the rubrics were revised based on the pilot results, these students are not included in the main study.

To sum up, three frameworks of rubrics were developed, covering acting democratically, acting socially responsible and dealing with differences respectively. Each framework consists of three dimensions (knowledge, attitude and skills) and four levels (A, B, C, D). Each cell in the frameworks contains a general description and three examples of the corresponding knowledge, attitude and skills respectively. The cell representing the attitude towards dealing with differences at level B is shown in Figure 1 as an example. The complete frameworks can be found in Appendix A.

	Level A	Level B	Level C	Level D
Knowledge	…	…	…	…
Attitude (I believe…)	…	You assume everyone should abide by the same rules. You believe that while people differ, everyone is equal. For example: • I believe that men and women deserve equal opportunities. • I believe that people should be able to be who they want to be, as long as they abide by the rules. • I believe that it is normal to respect other religions, genders and sexual orientations.	…	…
Skills	…	…	…	…

Figure 1. Example description of dealing with differences – attitude – level B.

Respondents

For the main study, we collaborated with thirteen civics teachers in fourth-year general secondary education (grade 10) and six career and citizenship teachers in first-year vocational tertiary education (grade 11). These teachers incorporated either the self-assessment or the reflection assignment (see next sections) into their lessons at some point between September 2016 and February 2017. One of the three rubrics was selected by the teacher based on the topic students would be working on during that lesson or period. For example, one of the common themes in civics classes is 'the diverse society', which means the teacher would select the framework on dealing with differences. All students had to complete the assignment as part of the regular lessons.[4] A total of 716 students took part in the study (Table 1).

Table 1. Data description.

School level	N	Mean age	% female	Self-assessment	Reflection assignment
Secondary	456	15.4	50%	383	73
Vocational	260	17.4	57%	218	42
Total	716	16.2	53%	601	115

---

[4] Students were informed that the assignment was also part of a research project, and that they could withdraw from participating in the study without consequences. Twenty-two students indicated that they did not want to participate.

Self-assessment

Self-assessment enables students to get a better understanding of the capabilities they need for learning (Fraile, Panadero, & Pardo, 2017; Kostons, Van Gog, & Paas, 2012). By evaluating how students assess their own competences by using the frameworks, we aim to learn if and how rubrics can help students to evaluate their own development. The self-assessment assignment was completed by 601 students. Students assessed their competences on one of the three social tasks: 135 students completed the assignment for acting democratically, 172 students for acting socially responsibly and 294 students for dealing with differences. The assignment was made in class, and students generally completed the self-assessment within half an hour. Students were presented with the framework and instructed to answer three questions concerning their knowledge, attitude and skills: 'What level best describes you?'; 'Why?' and 'Can you give an example?'. Students were given an answer sheet on which to select their level on each dimension and provide their explanations and examples. To help students provide relevant answers, they were prompted with 'I know…', 'I believe…', and 'I can…' for knowledge, attitude and skills respectively.

The explanations students provided to the second and third questions were assessed on two criteria: 1) is the explanation relevant to the social task; and, if so, 2) is the explanation adequate to the selected level? The first question was intended to assess whether students understood the assignment and the topic. The second question was intended to assess whether students were able to accurately assess their own citizenship competences using the framework and provide sufficient justification. Students' explanations were graded by the first author and three trained assessors. The explanations by the first 40 students were assessed by all four assessors, after which differences in grading were discussed to reach a consensus.[5] The other explanations were assessed by three assessors: the first author and two of the other assessors in rotation. Final grading was based on simple majority of the three assessors.

Reflection assignment

We also tested whether rubrics could be used to assess students' answers on a reflection assignment. The reflection assignment was completed by 115 students. Each student completed the assignment on one of two social tasks: 42 students on acting socially responsible and 73 students on dealing with differences (see

---

[5] Both the first 40 students and the other students are included in the analysis.

Appendix C for the assignments). We developed two assignments which followed a lesson or series of lessons on the topic. The reflection assignment on acting socially responsible focused on perception of social inequalities. One of the questions was 'What positive consequences can social differences have?'. The reflection assignment on dealing with differences focused on cultural differences and prejudices about immigrants. One of the questions was 'Are there prejudices about this group of immigrants in the Netherlands?'.

For both assignments students' answers were graded A, B, C, or D on the knowledge and attitude dimensions of the corresponding framework. The first class of 21 students were graded by all four assessors, and differences in grading were discussed.[6] The remaining students were assessed by three assessors: by the first author and by two of the assessors in rotation. If at least two assessors had scored the same level, the student was assigned to that level. If all three assessors gave a different score, no level was assigned.

Comparing rubrics and questionnaires

The majority of instruments assessing citizenship competences uses a combination of a test and questionnaire. Because we intend to develop a new instrument, the evaluation of its suitability and validity also leads to a comparison between rubrics on the one hand, and tests and questionnaires on the other (cf. Reddy & Andrade, 2010). We are interested in comparing the similarity of student outcomes based on both instruments , i.e. convergent validity. To this end, we used an instrument based on multiple-choice questionnaires measuring knowledge, attitudes and skills on dealing with differences. To measure attitudes and skills, we used the aforementioned Citizenship Competences Questionnaire (CCQ) developed by Ten Dam et al. (2011); to measure knowledge, we used a citizenship knowledge test based on the same framework. A total of 108 students completed the test and questionnaire on knowledge, attitudes and skills relevant to dealing with differences. Of this group, 89 students had self-assessed themselves using the rubrics, while 19 had completed the reflection assignment.

Results

Self-assessment

To evaluate how suitable our instrument is for self-assessment of citizenship competences we are interested in four aspects. To what extent:

---

[6] Both the first 21 students and the other students are included in the analysis.

- are the students able to assign themselves to one of the four levels of knowledge, attitude and skills?
- are the explanations provided by the students relevant to the social task?
- are the explanations provided by the students adequate for the level they selected?
- do different assessors agree on the answers to these last two questions?

Starting with whether students are able assess their own competences, 97 to 100 percent of students assigned themselves a level of knowledge, attitudes and skills. Students who selected two adjacent levels were considered to be in the lower of the two; since the levels are incremental, these students were considered 'on their way to the higher level, but not quite there yet'. The remaining 0 to 3 percent of students failed to make a clear choice. Figure 2 shows that most students assessed themselves to be at level B or C on all three dimensions.

For those students who selected a level, we evaluated whether the explanation they presented was relevant to the dimension (knowledge, attitude and skill) and the social task. In order for an explanation to be considered relevant, it had to relate to both the social task and the dimension. Table 2 and Figure 2 show that the vast majority of students were able to provide a relevant explanation. The percentages are similar for all three dimensions, with the highest scores for attitude on all social tasks. The percentages are somewhat lower for skills, which shows that the students had slightly more difficulty providing a relevant explanation, regardless of the social task. The low percentage of relevant explanations for knowledge about acting socially responsible shows a deviation from the pattern and appeared to be caused by students explaining their attitude rather than their knowledge.

Table 2 and Figure 2 also show the percentage of students who provided an adequate explanation for the level they selected. An explanation could not be adequate without being relevant. For an explanation to be adequate, the student had to provide an explanation that reflected the proficiency of the selected level without simply copying the contents of the framework. Here we see considerable differences from the results for relevance. About half the students provided an adequate explanation for their attitude, and close to one in four for their knowledge and skills. Figure 2 shows that the percentage of students providing an adequate explanation is also highly related to the selected proficiency level. Two reasons for not providing an adequate explanation appeared most prominent: some students overestimated themselves, providing explanation that were more adequate to a lower level; and many students stayed too close to the text in the framework,

simply copying parts of it, so that their explanation, although relevant, could not be considered adequate.
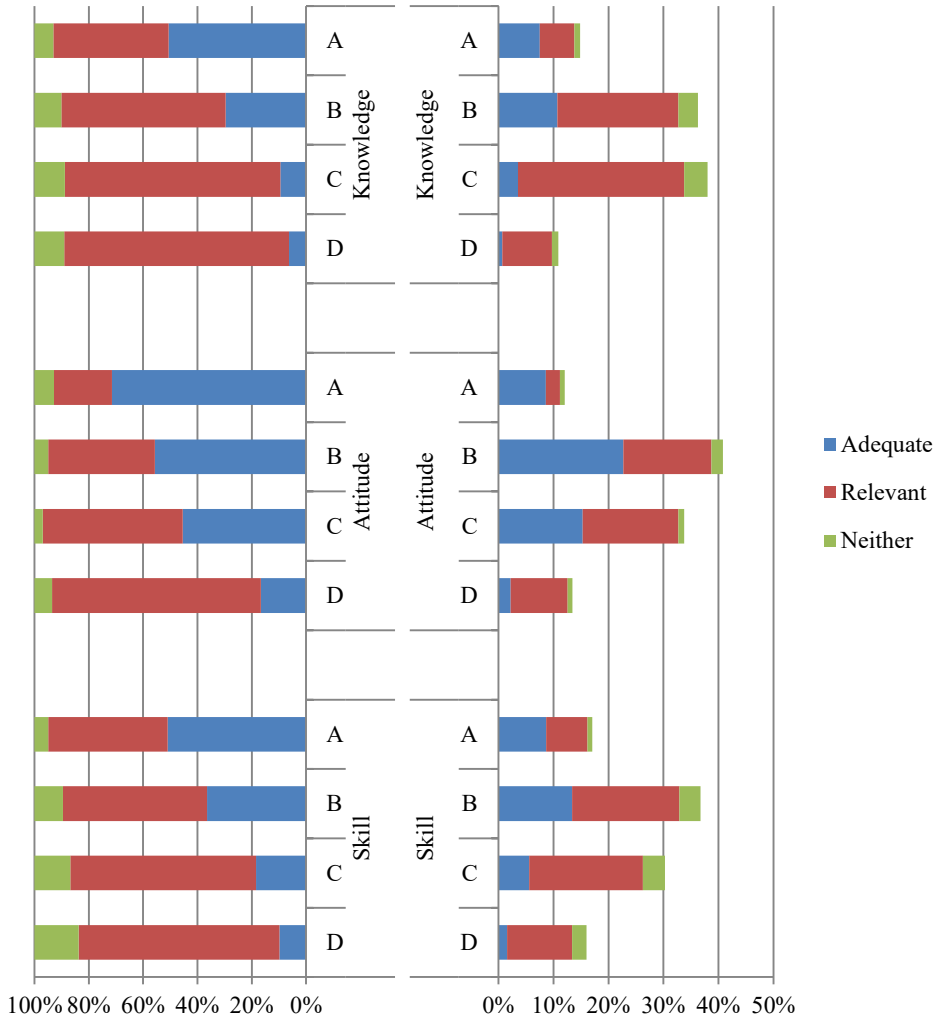

Figure 2. Results from student self-assessment.

Note: The graph on the left shows the proportion of adequate (and thus relevant), relevant (but not adequate), and neither relevant nor adequate explanations for each level. The graph on the right shows the distribution of students over the levels.

Table 2. Percentage of relevant and sufficient explanations per social task.

Social task	Knowledge (N=600)		Attitude (N=591)		Skill (N=590)	
	Relevant	Adequate	Relevant	Adequate	Relevant	Adequate
Acting democratically	91%	29%	92%	42%	88%	28%
Acting socially responsible	78%	24%	94%	61%	81%	28%
Dealing with differences	92%	17%	94%	44%	89%	29%
Total	88%	22%	94%	48%	86%	28%

The results are based on majority agreement between assessors. Inter-rater reliability of rubrics is commonly expressed as the percentage of agreement between assessors (Jonsson & Svingby, 2007). Table 3 shows the agreement between assessors in this study. Average pairwise agreement shows the likelihood that any two of the assessors agree. Agreement between pairs of assessors was around 90 percent for relevance and around 80 percent for adequacy. Agreement between all three or four assessors (i.e. full agreement) was 85 to 89 percent for relevance and 65 to 76 percent for adequacy. Because inter-rater agreement does not correct for chance, we also calculated Cohen's Kappa for each pair of assessors. Cohen's Kappa ranged from 0.164 to 0.838, with a mean around 0.5 for all measures. In their review of scoring rubrics, Jonsson and Svingby (2007) report Kappa values between 0.20 and 0.63. Although benchmarks for the size of Cohen's Kappa are essentially arbitrary, 0.4 to 0.6 can be considered moderate strength of agreement (Landis & Koch, 1977). All mean Cohen's Kappa's in our study are within this range.

Table 3. Interrater agreement for self-assessment.

	Knowledge		Attitude		Skill	
	Relevant	Adequate	Relevant	Adequate	Relevant	Adequate
Mean pairwise agreement	90%	81%	93%	76%	92%	83%
Full agreement	85%	74%	89%	65%	89%	76%
Mean of Cohen's Kappa	0.496	0.506	0.478	0.512	0.578	0.589

Overall, the results of the self-assessment show that nearly all students are able to assign themselves to a level, and nine out of ten are able to provide a relevant explanation. Providing an adequate explanation for the level selected appeared considerably more difficult. About half the students are able to provide an adequate explanation for their attitude but only one in four for knowledge and skills. As might be expected, providing an adequate explanation becomes more difficult as the level selected is higher. Although inter-rater agreement is high overall, it is inflated by chance. After controlling for chance, inter-rater agreement

is generally moderate, which means assessors appear to moderately agree on how to evaluate the students' explanations.

Reflection assignment

To evaluate the usefulness of our instrument for assessing students' reflections on a social task we are interested in two aspect. To what extent:

- can students be assigned a level of knowledge and attitude?
- do different assessors agree on students' level of proficiency?

A total of 115 students in five classes completed a reflection assignment. All students could be assigned to a level for attitude and 97 percent for knowledge. Figure 3 shows the levels to which students were assigned and the extent to which the assessors were in agreement. Most students scored A or B on knowledge and B on attitude.

Table 4 shows the extent of agreement between assessors. Agreement between pairs of assessors was around 70 percent for knowledge and around 80 percent for attitude. All assessors agreed in 59 percent of cases for knowledge, and 70 percent of cases for attitude. Cohen's Kappa ranged from 0.408 to 0.758, with means of around 0.5 for knowledge and 0.6 for attitude. Cohen's Kappa of 0.4 to 0.6 can be considered to reflect moderate strength of agreement and 0.6 to 0.8 can be considered substantial agreement (Landis & Koch, 1977).

Table 4. Interrater agreement between assessors for the reflection assignment.

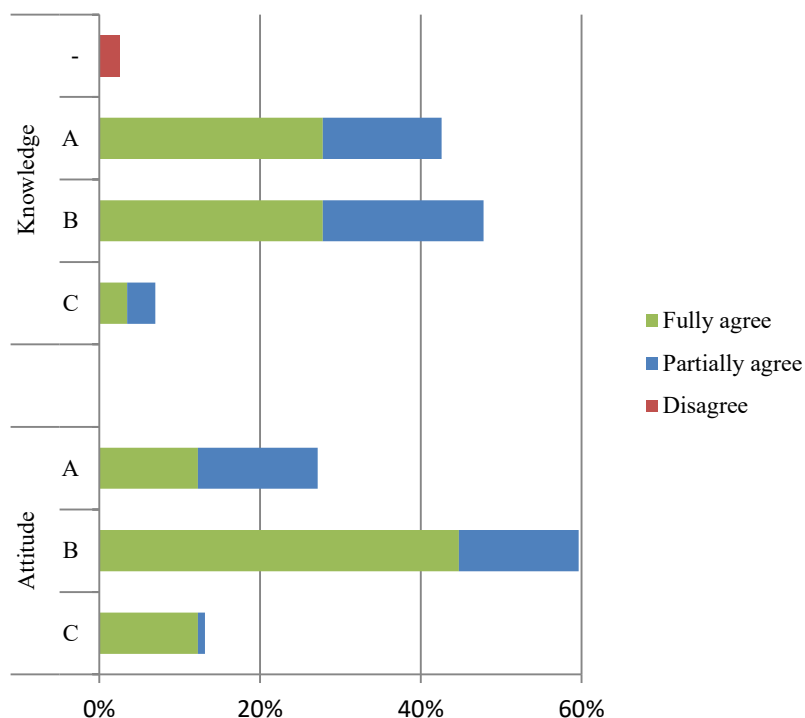|	Knowledge	Attitude
Mean pairwise agreement	72%	79%
Full agreement	59%	70%
Mean of Cohen's Kappa	0.534	0.618

Figure 3. Results of the reflection assignment.

Note: The graph shows the distribution of students over the levels. The colours show whether all assessors agreed on the assigned level.

Overall, the results of the reflection assignment show that nearly all students could be assigned to a level based on simple majority agreement between the assessors. The large agreement between assessors can in part be explained by chance; however, after controlling for chance, agreement remains moderate to substantial. The distribution of scores is somewhat skewed towards the lower end, with most students scoring A or B. This distribution is similar to that found for students who provided an adequate explanation on the self-assessment assignment. The rubrics appear reasonably well suited for assessing the students' knowledge and attitude when reflecting on a social task.

Comparing rubrics and questionnaires

The previous sections show that rubrics, like tests and questionnaires, can be used to assess students' citizenship competences. This raises the question to what extent assessment using rubrics for citizenship competences produces similar outcomes to tests and questionnaires. We asked 106 students to complete a short subset of items

taken from the Citizenship Competences Questionnaire (CCQ) on their knowledge, attitudes and skills on dealing with differences. Of these 106 students, 87 had completed the self-assessment and 19 had completed the reflection assignment. Table 5 shows the number of students for whom we could compare scores.

Table 5. Number of students per component for self-assessment and reflection assignment.

Dimension	Self-assessment			Reflection
	All	Relevant	Adequate	
Knowledge	76a	72	8	19
Attitude	87	85	28	19
Skills	86	81	19	-b

Notes: a) Thirteen students failed to notice that the knowledge items were on the back of the form, and only completed the attitude and skills items. b) Because the reflection assignment only assessed knowledge and attitude, no score for skills was given.

Figure 4 shows the boxplots for the relationship between the CCQ and rubrics using self-assessment (left) and the reflection assignment (right). All three graphs for self-assessment show a slight upward trend, but most notably a large overlap between the students' results at all four levels. The graphs for the reflection assignment in relation to the CCQ show near equal scores on the questionnaire for students at all rubric levels. Analysis of variance confirms no significant relation between the two instruments for either self-assessment or the reflection assignment.[7] We also see some indication of a ceiling effect for knowledge and skills on the CCQ. These analyses are based on a small sample, and a larger sample is needed to make any robust claims about a relationship between the two instruments.

---

[7] Five analyses of variance were conducted, each of which showed non-significant relations between scores both variables:
Knowledge self-assessment – CCQ [$F(3, 72) = 1.92$, $p = 0.131$];
Attitude self-assessment – CCQ [$F(3, 83) = 1,797$, $p = 0.154$];
Skills self-assessment – CCQ [$F(3, 82) = 1.21$, $p = 0.312$];
Knowledge reflection assignment – CCQ [$F(2, 16) = 0.06$, $p = 0.947$];
Attitude reflection assignment – CCQ [$F(1, 17) = 0.64$, $p = 0.435$].

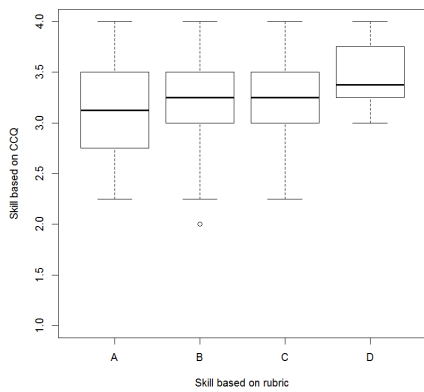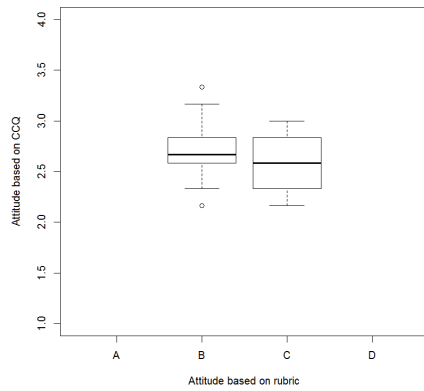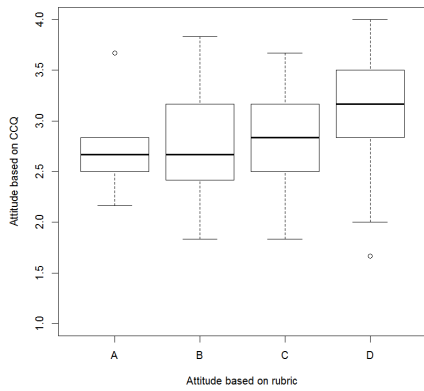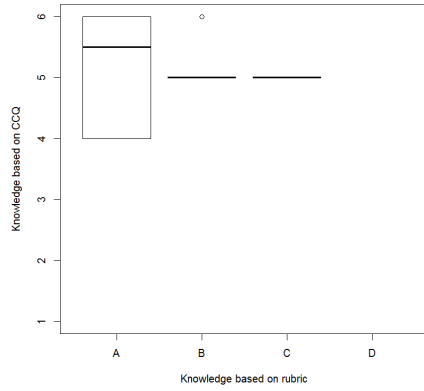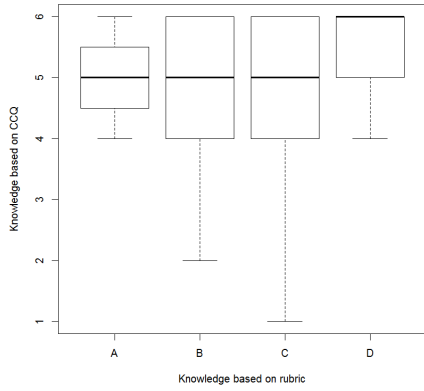Self-assessment   Reflection assignment

Figure 4. Relationship Rubrics – CCQ questionnaire: dealing with differences.

Conclusion

The aim of this study was to evaluate if and how rubrics can support assessment of students' citizenship competences. We developed and tested three rubrics of social tasks relating to aspects of citizenship relevant to young people. While rubrics are considered suitable for assessing complex competences (Jonsson & Svingby, 2007), no earlier studies have been conducted into their application for assessing citizenship competences. Overall, the rubrics we developed showed several strengths and weaknesses when used for both self-assessment and assessment of students' work on a reflection assignment.

The results of using rubrics for self-assessment show that the vast majority of students are able to assign themselves to one of the levels and to provide a relevant explanation for their self-assessment. This supports the conclusion that students understand the topics at which the rubrics are aimed and can assess their own competences. The assignment gave students the opportunity to explain the relevance of the social task to their own lives, which also allowed us to assess to what extent students are able to justify their self-assessed level. Agreement between the assessors whether an explanation was relevant and adequate was generally of moderate strength. The majority of students assessed themselves at level B or C, but after checking for sufficient explanations most scored A or B. The explanations students provided for the level they chose often left something to be desired. For attitude, about half of the students were able to provide an explanation that could be considered adequate for the level they had chosen. For knowledge and skills, this applied to about one in four students. Students appeared to have less difficulty explaining what they think about a social task than explaining what they know or are able to do. The large number of inadequate explanations is probably due to two reasons: students overestimating their own competences, and inadequate ability or motivation to provide grounds for their choice. While it is impossible to distinguish between these reasons based on the data, the general quality of students' explanations was rather low, with students often copying parts of the framework. Rubrics are often used in post-secondary education and the students in our study were probably unfamiliar with this type of task. The cognitive load which the assignment placed on the students' reflection and writing skills may therefore have played a part too, and several teachers reported they felt they needed to support students to complete the self-assessment task. Self-assessment of citizenship competences using rubrics can therefore be considered more suitable for citizenship attitudes than for knowledge and skills, and students need support in self-assessing – and particularly explaining – their competences.

When used for assessment of a reflection assignment, the rubrics allowed us to assign students to a level of knowledge and attitude with moderate to substantial reliability. The results showed that students could be assigned to a level of knowledge and attitude by all assessors, and in 97 percent of cases at least two of the three assessors agreed on this level. The distribution of scores was skewed towards the lower levels, with most students scoring A or B. We saw large variation in the quality of students' answers, where students who scored higher (i.e. C) often provided more elaborate answers. Overall, the rubrics appeared to be suitable for assessing student knowledge and attitude, and the preparation of assessors on how to apply the rubrics to students' work produced moderately to substantially reliable results.

Agreement between assessors on both assessment approaches was generally of moderate to substantial strength. Since one of the aims of assessing citizenship competences through rubrics was to provide room for personally meaningful conceptions of citizenship, a moderate agreement might be considered acceptable for low-stakes purposes. Asking students to explain their views did allow us to uncover students' misconceptions, something which is generally not possible with Likert-type items. The lack of a clear-cut relationship between outcomes for the rubrics and the questionnaire does call for further investigation of the effect of the method used for assessing the outcomes for students' citizenship competences. Both the self-assessment and reflection assignment depended on the quality of student input. Further study is required to answer the question whether the results of this type of citizenship competence measurement depends (too much) on the students' reflection and writing skills. Perhaps students who are more used to self-assessment can provide clearer and more elaborate answers. One of the challenges appears to be to teach and motivate students to explain themselves more fully. Future research is also necessary to investigate if (self-) assessment of citizenship competences using rubrics leads to improved future learning.

Discussion

Reliability and validity

Reliability and validity are essential to any assessment instrument, and studies using rubrics should strive to provide a clear explanation of the way in which these were established (Reddy & Andrade, 2010). The Methods section describes how the rubrics we developed are based on an existing theoretical framework of citizenship competences relevant to the lives of Dutch students as young citizens (Ten Dam et al., 2011). To improve the quality and validity of these rubrics, we

conducted several pilot studies in which teachers, students and researchers were consulted. Although this does not guarantee validity, it underscores the rigor that has gone into their development. To gain an understanding of the reliability of our rubrics, all student material was assessed by three assessors. Inter-rater reliability is of interest because rubrics are often applied differently by different assessors (Rezaei & Lovorn, 2010). Overall, we found moderate to substantial agreement between assessors, which is in line with other studies (Jonsson & Svingby, 2007). Assessors were trained by providing explanations of the rubric structures before they used them, and the assessment of the first students was discussed with the assessors. We noticed little to no increase in agreement after discussing the first results. To reach higher agreement, the assessors could perhaps have collaborated more intensively (cf. Meier, Rich, & Cady, 2006).

Reflection and writing skills

To investigate the convergent validity, 106 students were assessed on dealing with differences using both rubrics and a test and questionnaire. Although the lack of a significant relationship between either rubric approaches and the Citizenship Competences Questionnaire can be explained by the limited sample size, studies focusing on other topics also find modest correlations between rubrics and other measures or instruments (Jonsson & Svingby, 2007). The methods used to assess competences appear to affect the estimation of student performance. In more open-ended approaches such as rubrics, the students' reflection skills may influence the quality of their work. The quality of student writing has also been found to influence assessor judgements (Razaei & Lovorn, 2010). Both the self-assessment and the reflection assignment relied on the students' reflection and writing skills. In a general sense, we were rather underwhelmed by the quality of student writing, although there were definitely exceptions. It is possible that students' reflection and writing skills affected their performance. Motivation may have also played a part. The assessment was low-stakes for students, which may have negatively affected their motivation to make an effort (cf. Richardson, 2010).

Effect on learning

Although rubrics are considered promising for supporting student learning, the evidence for this is scarce (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). It is unclear which activities are required for rubrics to influence learning. Most studies using rubrics to support assessment for learning investigate the effects of using rubrics for self-assessment (Panadero & Jonsson, 2013). In this study, we therefore also focused on self-assessment as a necessary – but not sufficient –

condition for assessment for learning. We are however reluctant to make any claims about the benefits of using rubrics for students' development of citizenship competences, especially since merely providing students with a rubric is unlikely to have any effect without teachers actively involving students (Fraile et al., 2017; Reddy & Andrade, 2010). Moreover, there is a lack of longitudinal research into mechanisms affecting the development of student citizenship competences (cf. Amnå, Ekström, Kerr, & Stattin, 2009). Additional research into the development of citizenship competences and mechanisms for improving student learning is necessary to identify effective methods of promoting the development of citizenship competences – including the use of rubrics.

CHAPTER 4: ASSESSING CITIZENSHIP ATTITUDES USING RUBRICS[1]

Abstract

Schools' role in the development of students' attitudes towards society is increasing deliberated, which also leads to question how these attitudes can be assessed. Assessment instruments typically use Likert-type items, to measure the extent to which students agree or disagree to certain statements. In this study we aim to evaluate an alternative approach to assessing citizenship attitudes: using rubrics. The theoretical framework describes how we conceptualized attitudes towards acting democratically, acting socially responsible, and dealing with differences at four levels. Students in Dutch general secondary and vocational tertiary education were asked to assess which level they felt best described their attitudes, and to explain why. The results show students are generally implicitly supportive of democratic principles, show a willingness to help others, and respect people's differences. We find several differences between students' attitudes based on their background. The implications of using a comprehensive assessment of citizenship attitudes are further discussed.

Introduction

Education is widely considered to contribute to students' development of the competences necessary to act as citizens in democratic societies (cf. Friedman, 1955; Nie, Junn, & Stehlik-Barry, 1997). The development of these competences concerns *inter alia* teaching ways of democratic decision-making and workings of laws and institutions, developing the skills to participate in democratic processes, and developing democratic attitudes. The latter is the topic of this paper. The democratic attitudes of citizens – hereafter referred to as citizenship attitudes – pertain to how citizens feel about their rights and responsibilities in relation to others and society at large. Citizenship attitudes are relevant to both the collective and the individual (cf. Rychen & Salganik, 2003). They affect the collective by helping to sustain democratic societies by supporting democratic ideas, values and practices (cf. Dahl, 2000; Eckstein, 1966; Welzel, 2007) and they affect the individual's ability to act effectively as socially competent citizens in a democratic society (Ten Dam & Volman, 2007).

The term attitudes is used widely in research, and commonly refers to positive or negative feelings about concrete or abstract ideas, persons, situations or

---

[1] Based on: Daas, R., Dijkstra, A. B., & Karsten, S. (submitted). Assessing citizenship attitudes using rubrics.

events (Allport, 1935). We distinguish attitudes from values, with attitudes being personal rather than collective. We further consider attitudes to be a holistic concept, encompassing cognition, affect and connotation and composed of specific beliefs, feelings and action tendencies (Ajzen, 2005). More specifically, citizenship attitudes refer to such attitudes as the willingness to engage in democratic processes but also to views of social justice (cf. Council of Europe, 2016). These not only concern attitudes towards societal issues such as immigration but more broadly towards 'living together' in a democratic society.

Adolescence is generally regarded as a period of 'impressionable years', in which young people's citizenship attitudes are more susceptible to change than in later life. Various studies have shown that citizenship attitudes developed in adolescence grow more stable over the course of a person's life (Abdelzadeh & Lundberg, 2017; Alwin & Krosnick, 1991; Hooghe, Dassonneville, & Marien, 2015; Russo & Stattin, 2017). However, we still know little about how students' citizenship attitudes develop during adolescence. For example, it is not clear how students develop support for democratic attitudes during secondary school (Geboers et al., 2015; Keating et al., 2010). Students' background characteristics also appear to play a role. Students' citizenship attitudes have been shown to differ with respect to gender, age, cognitive level, socioeconomic status, and ethnic background (Geijsel et al., 2012; Schulz et al., 2010, 2017).

The development of students' citizenship attitudes is generally assessed using surveys with multiple-choice questionnaires (e.g. Keating et al., 2010; Schulz et al., 2010, 2017). This type of assessment has proven very effective in studies involving large groups of students to provide a standardized measure of their citizenship attitudes. Students are typically asked to indicate the extent to which they agree or disagree with certain statements. However, they are generally not asked why they make a certain choice. This means that their underlying beliefs or arguments are difficult to assess with such instruments (see Chapter 5; cf. Duckworth & Yeager, 2015). Rubrics could potentially offer a more development oriented assessment approach (cf. Black & William, 1998).

A rubric is essentially a matrix consisting of two dimensions: the assessment criteria and the levels of proficiency. Each cell in this matrix specifies a level of expected performance. Rubrics are increasingly common in educational assessment (cf. Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). However, this approach has rarely been used to measure citizenship competences (see Chapter 3). In general, rubrics are used to clarify learning goals, communicate those goals to students, act as guidelines to provide progress feedback to students and assess learning results (Andrade, 2005). There are no clear-cut theoretical

limits to the dimensions a rubric may contain; however, the number of proficiency levels is typically three to five. The number of assessment criteria can range from a single criterion to over twenty.

In this study we aim at a comprehensive method for assessing students' citizenship attitudes by using a rubrics based approach that provides opportunities for students to include information about their personal views and experiences with citizenship attitudes. We ask students not only to assess their level of citizenship attitudes, but also to explain why they feel that level in the rubric best describes them. This way students' explanations not only allow us insights into their self-assessed level of citizenship attitudes, but also to identify dissonance between the indicated level and the explanation they provide. Since earlier studies assessing citizenship competences have shown that students' citizenship attitudes vary with their background, we will also investigate whether these differences are also found when assessing citizenship attitudes using rubrics. The aim of this study is to evaluate what assessment using rubrics can tell us about students' citizenship attitudes. To this end we aim to answer three questions: 1) *What level do adolescents assess their citizenship attitudes to be?*, 2) *To what extent are they able to adequately explain their citizenship attitudes?* and 3) *To what extent do the citizenship attitudes of adolescents relate to their individual background?* The next section will describe how we developed three rubrics, each describing four levels of citizenship attitudes.

Theoretical framework

To develop rubrics for the assessment of citizenship competences we use the framework developed by Ten Dam et al. (2011). These authors developed a measurement instrument for citizenship competences by considering four 'social tasks'. The authors refer to social tasks as activities that students are likely to perform as young citizens. We have selected three of these: acting democratically, acting socially responsible and dealing with differences. With this conceptualization of citizenship competences, and more specifically the citizenship attitudes of students, we link up with a tradition in which citizenship is not only seen as relevant to the students' future role as citizens but is also regarded as relevant to the everyday lives of young people and visible in the situations they may likely encounter (cf. Rychen & Salganik, 2003; Ten Dam et al., 2011). The premise is that development of citizenship and citizenship attitudes takes place in social contexts.

Effective performance of social tasks requires relevant knowledge, skills and attitudes. To assess students' citizenship competences we developed three rubrics, each describing four levels of knowledge, skills and attitudes (see Chapter 3). In this study, we will focus on one dimension – attitudes. Students can, for example, become more sensitive to the social and cultural differences among the people around them and develop more complex perspectives on how to deal with these differences.

The attitudes for each social task are described in levels A through D. As will be discussed more fully in the next section, level A describes attitudes that can be considered legitimate, but assumes living together in society – the focus of the investigated attitudes – as unproblematic. Level B reflects what can be considered attitudes that are implicitly supportive of democratic principles. Level C describes attitudes that can be considered explicitly and deliberately supportive of democratic principles. Level D describes attitudes that are sensitive to complex situations, and supportive of promoting democratic attitudes among others. In an educational setting, moreover, particularly levels A and D could signal to the teacher and students that they need more support or greater challenges respectively. The complete rubrics are included in Appendix A (translated from Dutch).

Acting democratically

Democratic citizenship is reflected in daily life in the way people act democratically when making decisions. A robust democracy means all citizens are involved in and participate in decision-making. According to Ten Dam et al. (2011, p. 357) "a young person with such attitudes wants to hear everyone's voice, enter into a dialogue and make an active, critical contribution." Attitudes towards acting democratically thus do not refer to wanting to make the best decision, but rather to wanting to make the decision in the best possible way (cf. Barber, 1984). This therefore not only involves wanting to actively engage in decision-making, but also attitudes such as being sensitive to minority viewpoints and interests. Students can develop multiple considerations to take into account when making decisions, such as being sensitive to differing contexts and choosing to find consensus, vote or deliberate (Nieuwelink, Dekker, Geijsel, & Ten Dam, 2016). The descriptions of attitudes towards acting democratically therefore focus on considerations that students take into account in the decision-making process, where students with a more developed attitude are more sensitive and willing to take more considerations into account.

Level A describes an attitude that regards democratic decision-making as instrumental and unproblematic. Rules, once agreed upon, can be considered

permanent, and convictions about decision-making are susceptible to the 'tyranny of the majority'. These students take (a limited selection of) democratic principles into consideration separately. Level B can be considered preferable to level A by describing an attitude that supports the inherent value of acting democratically through support for democratic principles (e.g. equality and open debate). These students take multiple democratic principles into consideration. While this attitude supports democratic citizenship, it does not include an active interest in democratic processes. Level C builds on level B by describing a disposition to actively engage in democratic processes and a willingness to consider less-well represented viewpoints and interests. This level also emphasises that critical reflection on democratic decisions is important. Democratic processes are not only considered necessary to decision-making, but valuable in and of themselves. Level D is considered superior to level C in that it also advocates promoting an open democratic attitude in others and being sensitive to complex situations.

Acting socially responsible
Citizenship education is often envisioned to promote social cohesion and responsibility (cf. Council of Europe, 2010). Social responsibility stems from a commitment to contribute to the well-being of others and society for a greater good (Berman, 1990; cf. Rychen & Salganik, 2003). Acting socially responsible is aimed at promoting a sustainable way of co-existing with others and the environment. According to Ten Dam et al. (2011, p. 357) "a young person with such attitudes wants to uphold social justice, is prepared to provide care and assistance, does not want to harm another or the environment as a result of his or her behaviour." These attitudes underline that citizens in a democratic society share a common purpose (Barber, 1984). Students with a more developed attitude show more sensitivity to the wants and needs of others around them and to the principles underlining their actions, particularly reciprocity and sustainability (cf. Kohlberg, 1981).

Level A describes an attitude that considers social responsibility from the perspective of peaceful co-existence. It supports tolerance of others, but there is no explicit support for collaboration or common purpose. Rather, people should essentially not harm one another. Level B is distinguishable from level A because it describes an awareness of the value of collaboration and helping others. There is a sense that people need to contribute to a greater good. Level C builds on level B by including a more active desire to contribute to the lives of others and the environment. There is a sense of being part of a community in a general sense, and the need for balancing personal and common interests. Level D describes an

attitude that focuses not only on students' own attitudes but also on promoting a supportive attitude among others.

Dealing with differences

Democracy, by valuing individual freedom, presupposes plurality, that is, the existence of differences (cf. Van Gunsteren, 1998). Different viewpoints have to be accepted, represented and upheld for a democracy to function. By extension, differences between people have to be acknowledged, while fundamental human rights require that all people are considered equal. Democratic citizens develop attitudes on how they regard these different viewpoints, convictions, customs or ways of living. According to Ten Dam et al. (2011, p. 357) "a young person with such attitudes has a desire to learn other people's opinions and lifestyles, has a positive attitude towards differences." Some of the most significant contemporary debates in modern democracies pertain to issues involving ethnic, cultural, religious or sexual differences. As participants in a democratic society, young people are expected to develop attitudes towards dealing with these and other differences (cf. De Groot, 2011; Van Gunsteren, 1998). Students with a more developed attitude show more sensitivity to the backgrounds and implications of differences between people, and are more reflective of their attitudes towards others.

Level A describes an attitude that recognizes differences but essentially maintains that minorities should adapt to the majority. Level B focuses on a sense of equality: while people may differ, they are essentially equal. Differences should be allowed to exist within the boundaries of rules, but are generally considered as a deviation from the norm. Level C includes the sensitivity required to deal with differences and openness to those differences. This differs from the previous levels because differences are considered relevant in their own right. Level D describes an attitude that besides the sensitivity included in level C also includes the relevance of careful consideration of one's own attitudes in dealing with differences and a readiness to adapt these attitudes. Level D also again involves promoting these attitudes among others.

Differences in citizenship attitudes

Previous studies have shown citizenship attitudes differ based on students' background characteristics (Geijsel et al., 2012; Keating et al., 2010; Schulz et al., 2010, 2017). Several differences between students due to background characteristics emerge from these earlier studies. We pay specific attention to the studies by Munniksma et al. (2017) which like our study is based on data from

Dutch students, and Geijsel et al. (2012) which moreover uses the same theoretical framework by Ten Dam et al. (2011) as our study. Notably, all these studies use standardized surveys based on questionnaires (using Likert-type items) to assess citizenship attitudes.

Girls, by and large, score higher than boys on most measurements of citizenship attitudes. These differences are found in the International Civics and Citizenship Studies (ICCS; Schulz et al. 2010, 2017; Munniksma et al., 2017), the Citizenship Education Longitudinal Study (CELS; Keating et al., 2010) and in studies conducted in Dutch schools involving the three social tasks that are the focus of the present study (Geboers et al., 2015; Geijsel et al., 2012). In the study by Geijsel et al. (2012), girls scored the same as the boys on attitudes towards acting democratically and higher on attitudes towards acting socially responsible and dealing with differences.

The Dutch education system is highly tracked, with up to seven different tracks in general secondary education. These tracks are often aggregated to pre-vocational, pre-professional, and pre-university tracks. Students in higher educational tracks often score higher on measures of citizenship attitudes (Geijsel et al., 2012; Munniksma et al., 2017; Schulz et al., 2010, 2017). In the study by Geijsel et al. (2012), students in higher educational tracks scored higher on attitudes towards all three social tasks.

With respect to age differences, some earlier studies suggested that student citizenship attitudes develop non-linearly, with support for some attitudes increasing while support for others decreases (Geboers et al., 2015; Keating et al., 2010). Geijsel et al. (2012) found that older students in Dutch schools score similar to younger students on attitudes towards acting democratically and lower on attitudes towards acting socially responsible and dealing with differences.

Several studies found a relation between students' ethnic background and their citizenship attitudes, with students with an immigration background generally scoring higher depending on the attitude measured (Eckstein et al., 2015; Geijsel et al., 2012; Munniksma et al., 2017). In the study by Geijsel et al. (2012) students with an immigration background (based on their mother's country of birth) score higher on citizenship attitudes towards all three social tasks.

Methods

Data

The data was collected as part of a study conducted in the Netherlands in which we evaluated the feasibility of using rubrics to assess citizenship knowledge, attitudes,

and skills (see Chapter 3). The present paper focuses specifically on students' self-assessment of citizenship attitudes. Fourteen teachers with a total of 601 students in 28 classes collaborated in the study. All students self-assessed their attitudes for one of the three social tasks. Because not all students filled in the background questionnaire, we report the results for 567 students for whom the selected background characteristics are available. Teachers could choose which social task to assess depending on their curriculum, so the number of students for each social task is unevenly distributed: 130 for acting democratically, 160 for acting socially responsible, and 277 for dealing with differences.

Students were either in the fourth year of general secondary education (64 percent) or the first year of tertiary vocational education (36 percent). Tertiary vocational education consists of four tracks. Our sample consists of students from the highest level of vocational education (*mbo-4*) and the pre-professional secondary education track (*havo*). While the groups we selected can be considered to represent the 'low' and 'high' tracks, the entrance requirements for students in these particular instances are the same.

To investigate differences due to student characteristics, we asked students to answer several background questions (see Table 1). Fifty-three percent of students were female. The proportion of females in vocational education is somewhat skewed between social tasks. Our 'acting democratically' sample included students attending the IT vocational education programme, and most students in that sector are male. Similarly, the 'dealing with differences' sample included two classes in the fashion programme that consisted of mostly female students.

The students in general secondary education were around 15 years old on average, while the students in vocational education were around 17 years old on average. This age difference could be expected because of the one-year grade difference between these groups, and it is common to find older students in vocational education. Because there are limited possibilities for analysing the effects of age differences in view of the high correlation between age and level of education, we dichotomized age to reflect relative age within educational level: in secondary education students up to 15-years-old were considered 'relatively young', in vocational education students up to 17-years-old were considered 'relatively young' (see Table 2).

To investigate the differences between students due to their ethnic background, we asked them in what country they were born and the native country of both their mother and their father. Five percent of students were born abroad. Of both fathers and mothers, 26 percent were born abroad. Foreign-born mothers came

from 45 different countries and foreign-born fathers from 40 countries, with about half having been born in Morocco, Surinam or Turkey. We classified students with either one or two foreign-born parents as having a migrant background, which constitutes 34 percent of our sample.

Table 1. Distribution of data according to student characteristics (N = 567).

	N	%
Social task		
Acting democratically	130	23%
Acting socially responsible	160	28%
Dealing with differences	277	49%
Educational level		
Secondary	363	64%
Vocational	204	36%
Gender		
Male	269	47%
Female	298	53%
Relative age		
Younger	355	63%
Older	209	37%
Country of birth parents		
Both Netherlands	348	66%
Other	176	34%

Table 2. Distribution of data according to student age (N = 564).

Educational level	Actual age								Relative age	
	14	15	16	17	18	19	20	21+	younger	older
Secondary	12	210	113	28	0	0	0	0	222	141
Vocational	0	3	59	71	31	18	7	12	133	68

Approach

In Dutch education, the promotion of citizenship is not limited to specific school subjects but is addressed school-wide. However, in general secondary education citizenship is (also) specifically discussed as part of *maatschappijleer* (civic studies). In vocational education, a similar subject is called *loopbaan en burgerschap* (career and citizenship). For this study, we collaborated with fourteen teachers of *maatschappijleer* in secondary education and *loopbaan en burgerschap* in vocational education. Each teacher chose what social task was to be assessed, based on the topics discussed in class during that period. For example, students

studying 'parliamentary democracy' (a mandatory theme in the civic studies curriculum) were assessed on acting democratically. Students were presented with the rubric by their teacher and asked to 1) indicate which level best described them, 2) explain why they believed this was so, and 3) give an example. To help them give relevant answers, students were prompted with the following cue: 'I believe… for example…'.

The explanations provided by the students were assessed to evaluate if the explanation provided was relevant to the social task and, if so, was adequate for the level selected. If both were the case, it seems plausible that the student is at the indicated level. The explanations were assessed by the first author and three trained assessors. Cohen's kappa for mean interrater agreement was close to 0.5, indicating moderate agreement (Landis & Koch, 1977). About half (48%) of the explanations were considered relevant and adequate (for more information, see Chapter 3).

Analyses

The structure of the data and the data collection do not allow for a single model of analysis. The 567 students included in our analyses only assessed their attitudes on one of three social tasks, which means our total sample consists of three subsamples. We have two dependent variables for each student: their self-assessed attitude level on one of the social tasks, and whether the explanation they provided was relevant and adequate. All variables used are either categorical or dichotomous. We thus use log-linear analysis for contingency tables to analyse the data.

Firstly, we investigate what level students assess their citizenship attitudes to be, irrespective of their personal background. Using a log-linear model without interaction effects allows us to estimate the odds of a choosing a particular level. We repeat this process for only those students who provided an adequate explanation with their self-assessed level. We use χ^2-tests to evaluate whether the distribution of students over the levels changes when comparing their self-assessed levels to only those supported by adequate explanations.

Secondly, we are interested to what extent citizenship attitudes of adolescents relate to their individual background. Due to the limited sample sizes and research methods, we could not include all variables for each of the three social tasks. We therefore built three models. The models were fitted both to the self-assessed level and to only the data from students who provided an adequate explanation. The independent variables of interest are the educational level, gender, age and ethnic background of the students. The number of cells in a full contingency table (i.e. the dimension) is Attitudes (4) x Educational level (2) x

Gender (2) x Age (2) x Ethnic background (2) = 64. Table 3 shows the variables included in the models for each social task. Due to the relatively small sample size compared to the number of cells in the contingency tables, several cells contain a zero count. These pose a problem when considering complex interactions in log-linear models (Clogg & Eliason, 1987). We therefore only included baseline (single factor) effects and two-factor interactions in the models.

Table 3. Variables included for analyses.

	Acting democratically	Acting socially responsible	Dealing with differences
Educational level	+		+
Gender	+	+	+
Age		+	+
Migrant background		+	+

Results

Acting democratically

A total of 130 students assessed their attitudes towards acting democratically, 55 of whom provided an adequate explanation to the level they selected. Figure 1 shows the distribution over the levels for both groups, and Table 4 shows the results of the log-linear analyses. When self-assessing their attitudes towards acting democratically, students are more likely to choose level B, marginally more likely to choose level C, and less likely to choose level A or D. When considering only those students who provided an adequate explanation, students are more likely to choose level B or C, and less likely to choose level A or D. The distributions for self-assessed and adequately explained levels are unequal because the scores are lower when only those students who provided an adequate explanation are considered ($\chi^2 = 12.32$, df $= 3$, p < 0.01).
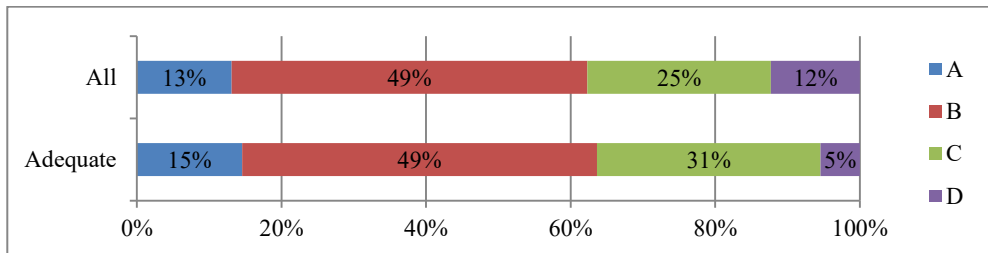

Figure 1. Distribution of students over levels acting democratically.

Table 4. Students' attitudes towards acting democratically.

	Self-assessed			Adequate explanations		
	coef.		odds ratios	coef.		odds ratios
(intercept)	1.86			0.91		
Attitude A	-0.48**	(0.20)	0.62	-0.25**	(0.31)	0.78
Attitude B	0.84***	(0.14)	2.32	0.97***	(0.23)	2.64
Attitude C	0.18	(0.16)	1.20	0.51**	(0.26)	1.66
Attitude D	-0.54***	(0.21)	0.58	-1.23***	(0.45)	0.29
Secondary	0.32***	(0.09)	1.37	0.24*	(0.14)	1.27
Boys	0.19**	(0.09)	1.21	0.02	(0.14)	1.02

Note: The coefficients are constrained to sum-to-zero. The italicized coefficients are redundant, but added for clarity. * significant at $p < .10$. ** significant at $p < .05$. *** significant at $p < .01$. Standard errors in parentheses.

We further investigate differences between students based on their background characteristics. The number of cells in the full contingency table (i.e. the dimension) is Attitudes (4) x Educational level (2) x Gender (2) = 16. We use the same model for the set of 55 students who provided a relevant and adequate explanation. Table 5 shows the results of both analyses.

When we first look at the levels students assessed themselves to be the results confirm that students are more likely to assess themselves at level B or C than level A or D. Boys are more likely than girls to choose level A, and marginally less likely to choose level B or C. The results do not show differences in attitudes between students enrolled in general secondary or vocational education. The sample includes a relatively large percentage of secondary school students and boys. Because the vocational education sample included a large proportion of boys (due to the IT students included), there is a negative relation for the proportion of boys in secondary education.

The results for the 55 students who provided an adequate explanation show that few effects are significant due to the small sample size. Here too, students are more likely to score level B or C than A or D. The only other significant relation is the proportion of boys in secondary education, again due to the sampled population in vocational education.

Table 5. Relations between student background and attitudes towards acting democratically.

	Self-assessed			Adequate explanations		
	coef.		odds ratios	coef.		odds ratios
(intercept)	1.40			0.41		
Attitude A	-1.00**	(0.44)	0.37	-0.52	(0.40)	0.60
Attitude B	1.07***	(0.20)	2.92	1.12***	(0.27)	3.07
Attitude C	0.38*	(0.22)	1.47	0.63**	(0.29)	1.88
Attitude D	-0.45*	(0.27)	0.63	-1.24**	(0.54)	0.29
Secondary	0.64***	(0.16)	1.89	0.35	(0.30)	1.42
Boys	0.74***	(0.23)	2.10	0.48	(0.31)	1.61
Secondary : Boys	-0.66***	(0.15)	0.52	-0.79***	(0.20)	0.45
Secondary : Attitude A	-0.09	(0.24)	0.91	-0.95	(0.71)	0.39
Secondary : Attitude B	-0.10	(0.16)	0.91	0.17	(0.35)	1.19
Secondary : Attitude C	-0.02	(0.19)	0.98	0.31	(0.36)	1.36
Secondary : Attitude D	0.21	(0.24)	1.23	0.48	(0.56)	1.61
Boys : Attitude A	0.94**	(0.45)	2.55	0.13	(0.70)	1.13
Boys : Attitude B	-0.36*	(0.20)	0.70	-0.34	(0.34)	0.71
Boys : Attitude C	-0.38*	(0.22)	0.68	-0.14	(0.35)	0.87
Boys : Attitude D	-0.20	(0.25)	0.82	0.35	(0.56)	1.42

Note: The coefficients are constrained to sum-to-zero. The italicized coefficients are redundant, but added for clarity. * significant at $p < .10$. ** significant at $p < .05$. *** significant at $p < .01$. Standard errors in parentheses.

Acting socially responsible

A total of 160 students assessed their attitudes towards acting socially responsible, 97 of whom provided an adequate explanation to the level they selected. Figure 2 shows the distribution over the levels for both groups, and Table 6 shows the results of the log-linear analyses. When self-assessing their attitudes towards acting socially responsible, students are more likely to choose level B, and less likely to choose level D. Of the 97 students who provided an adequate explanation only 1 scored level D. We therefore excluded this student from the log-linear analysis. When considering only those students who provided an adequate explanation, students are more likely to choose level B, and less likely to choose level C. The distributions for self-assessed and adequately explained levels are unequal because the scores are lower when only those students who provided an adequate explanation are considered ($\chi^2 = 59.40$, df = 3, $p < 0.01$).
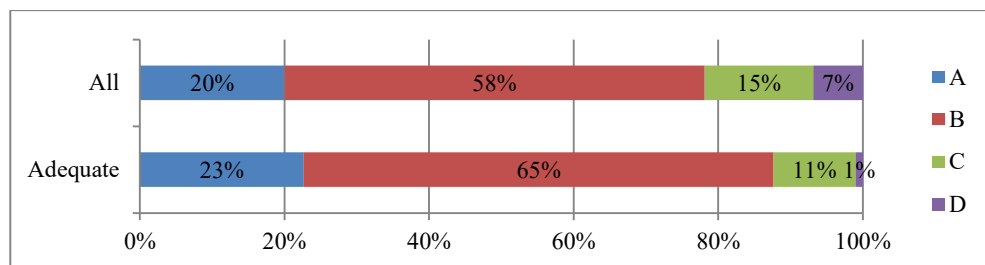
Figure 2. Distribution of students over levels acting socially responsible.

Table 6. Students' attitudes towards acting socially responsible.

	Self-assessed			Adequate explanations		
	coef.		odds ratios	coef.		odds ratios
(intercept)	1.27			1.08		
Attitude A	0.07	(0.16)	1.07	-0.15	(0.18)	0.86
Attitude B	1.14***	(0.13)	3.12	0.90***	(0.15)	2.47
Attitude C	-0.22	(0.18)	0.81	-0.75***	(0.21)	0.47
Attitude D	-1.00***	(0.24)	0.37			
Boys	-0.08	(0.08)	0.93	-0.09	(0.10)	0.91
Younger	0.26***	(0.08)	1.29	0.16	(0.10)	1.17
Dutch	0.16**	(0.08)	1.18	0.35***	(0.11)	1.43

Note: The coefficients are constrained to sum-to-zero. The italicized coefficients are redundant, but added for clarity. * significant at $p < .10$. ** significant at $p < .05$. *** significant at $p < .01$. Standard errors in parentheses.

To investigate differences between students based on their background characteristics the number of cells in the full contingency table (i.e. the dimension) is Attitudes (4) x Age (2) x Gender (2) x Migration background (2) = 32. Educational level is not included since all students were in secondary education. We use the same model for the students who provided an adequate explanation. Table 7 shows the results for both analyses.

The self-assessment results confirm that students are more likely to assess themselves at level B than the other levels. Boys are less likely than girls to score level C, and marginally more likely to score level A. Students of whom both parents are born in the Netherlands are marginally more likely to score level C. We found no differences between younger and older students. The sample included a relatively high percentage of younger students.

Table 7 also shows the results for the 96 students who provided an adequate explanation. In this case, too, students are more likely to score level B. The students who provided an adequate explanation included a relatively high percentage of students with a native Dutch background, which was not the case for

the total sample. In line with the results for the total sample, boys are more likely to score level A than girls.

Table 7. Relations between student background and attitudes towards acting socially responsible.

	Self-assessed			Adequate explanations		
	coef.		odds ratios	coef.		odds ratios
(intercept)	1.15			0.92		
Attitude A	0.16	(0.18)	1.18	-0.06	(0.23)	0.94
Attitude B	1.22***	(0.15)	3.40	1.02***	(0.20)	2.78
Attitude C	-0.39	(0.24)	0.68	-0.96***	(0.32)	0.38
Attitude D	-1.00***	(0.29)	0.37			
Boys	-0.00	(0.11)	1.00	-0.06	(0.15)	0.94
Younger	0.23**	(0.11)	1.26	0.11	(0.13)	1.12
Dutch	0.08	(0.12)	1.08	0.45***	(0.16)	1.56
Boys : Younger	-0.09	(0.08)	0.92	-0.17	(0.11)	0.84
Boys : Dutch	0.01	(0.09)	1.01	-0.04	(0.12)	0.96
Younger : Dutch	-0.05	(0.09)	0.95	-0.04	(0.12)	0.96
Boys : Attitude A	0.33*	(0.17)	1.40	0.55***	(0.20)	1.74
Boys : Attitude B	-0.11	(0.14)	0.89	-0.10	(0.16)	0.90
Boys : Attitude C	-0.52**	(0.21)	0.60	-0.45**	(0.23)	0.64
Boys : Attitude D	0.30	(0.26)	1.34			
Younger : Attitude A	-0.13	(0.17)	0.88	-0.12	(0.19)	0.89
Younger : Attitude B	0.07	(0.13)	1.07	0.12	(0.15)	1.12
Younger : Attitude C	0.20	(0.20)	1.22	0.00	(0.21)	1.00
Younger : Attitude D	-0.13	(0.25)	0.87			
Dutch : Attitude A	0.05	(0.17)	1.05	-0.25	(0.22)	0.78
Dutch : Attitude B	0.12	(0.14)	1.13	-0.10	(0.17)	0.91
Dutch : Attitude C	0.39*	(0.20)	1.48	0.35	(0.28)	1.41
Dutch : Attitude D	-0.57**	(0.26)	0.57			

Note: The coefficients are constrained to sum-to-zero. The italicized coefficients are redundant, but added for clarity. * significant at p < .10. ** significant at p < .05. *** significant at p < .01. Standard errors in parentheses.

Dealing with differences

A total of 277 students assessed their attitudes towards dealing with differences, 122 of whom provided an adequate explanation for the level they selected. Figure 3 shows the distribution over the levels for both groups, and Table 8 shows the results of the log-linear analyses. When self-assessing their attitudes towards dealing with differences, students are more likely to choose level B or C, and less likely to choose level A. When considering only those students who provided an adequate explanation, students are more likely to choose level B or C, and (marginally) less likely to choose level A or D. The distributions for self-assessed and adequately explained levels are unequal because the scores are lower when

only those students who provided an adequate explanation are considered ($\chi^2_{difference}$ = 60.41, df = 3, p < 0.01).

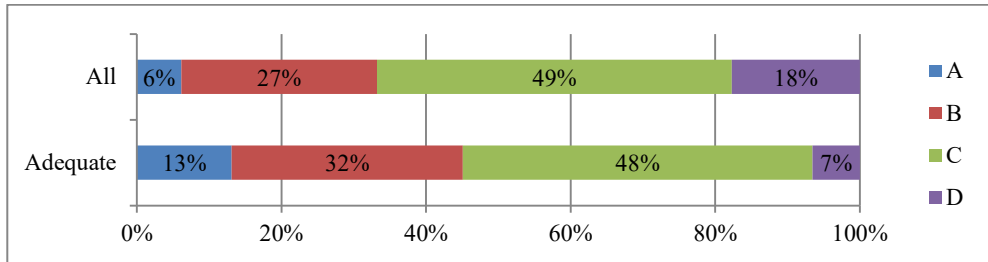**Figure 3. Distribution of students over levels dealing with differences.**

Table 8. Students' attitudes towards dealing with differences.

	Self-assessed			Adequate explanations		
	coef.		odds ratios	coef.		odds ratios
(intercept)	1.07			0.21		
Attitude A	-1.16***	(0.19)	0.31	-0.38*	(0.21)	0.69
Attitude B	0.33***	(0.11)	1.39	0.52***	(0.17)	1.67
Attitude C	0.92***	(0.10)	2.52	0.93***	(0.15)	2.53
Attitude D	-0.10	(0.13)	0.91	-1.07***	(0.28)	0.34
Secondary	-0.15**	(0.06)	0.86	-0.12	(0.09)	0.89
Boys	-0.15**	(0.06)	0.86	-0.30***	(0.10)	0.74
Younger	0.33***	(0.06)	1.38	0.25***	(0.09)	1.29
Dutch	0.39***	(0.06)	1.48	0.42***	(0.10)	1.52

Note: The coefficients are constrained to sum-to-zero. The italicized coefficients are redundant, but added for clarity. * significant at p < .10. ** significant at p < .05. *** significant at p < .01. Standard errors in parentheses.

The number of cells in the full contingency table (i.e. the dimension) is Attitudes (4) x Educational level (2) x Gender (2) x Age (2) x Migrant background (2) = 64. We use the same model for the set of 122 students who provided an adequate explanation. Table 9 shows the results for both analyses.

The self-assessment results confirm that students are more likely to assess themselves at level B or C than A or D. The interactions show no differences in attitudes according to level of education, gender, age or migrant background. Because the vocational education sample included two classes of fashion students (with mostly girls) and several classes of business services (with a relatively higher percentage of students with a migrant background; CBS, 2016), there are positive relations for the proportion of boys and students with Dutch parents in secondary education.

Table 9. Relations between student background and attitudes towards dealing with differences.

	Self-assessed			Adequate explanations		
	coef.		odds ratios	coef.		odds ratios
(intercept)	1.00			0.11		
Attitude A	-1.13***	(0.22)	0.32	-0.29	(0.26)	0.75
Attitude B	0.24*	(0.14)	1.27	0.39*	(0.22)	1.47
Attitude C	0.95***	(0.12)	2.58	1.03***	(0.18)	2.80
Attitude D	-0.06	(0.15)	0.95	-1.13***	(0.36)	0.32
Secondary	-0.22**	(0.09)	0.80	-0.21	(0.15)	0.81
Boys	-0.03	(0.09)	0.98	-0.23	(0.14)	0.79
Younger	0.23***	(0.09)	1.26	0.19	(0.13)	1.21
Dutch	0.41***	(0.09)	1.50	0.37***	(0.14)	1.45
Secondary : Boys	0.26***	(0.06)	1.29	0.23**	(0.10)	1.26
Secondary : Younger	-0.06	(0.07)	0.94	-0.11	(0.10)	0.90
Secondary : Dutch	0.13**	(0.07)	1.14	0.09	(0.11)	1.10
Boys : Younger	-0.09	(0.07)	0.92	-0.03	(0.10)	0.97
Boys : Dutch	0.03	(0.07)	1.03	-0.09	(0.11)	0.92
Younger : Dutch	0.03	(0.07)	1.03	-0.00	(0.11)	1.00
Secondary : Attitude A	-0.19	(0.20)	0.83	-0.27	(0.24)	0.77
Secondary : Attitude B	0.17	(0.12)	1.19	0.39**	(0.18)	1.47
Secondary : Attitude C	0.11	(0.11)	1.11	0.13	(0.17)	1.14
Secondary : Attitude D	-0.09	(0.14)	0.91	-0.25	(0.32)	0.78
Boys : Attitude A	0.29	(0.19)	1.34	0.47**	(0.23)	1.60
Boys : Attitude B	-0.11	(0.12)	0.90	-0.29	(0.19)	0.75
Boys : Attitude C	-0.11	(0.10)	0.90	0.02	(0.17)	1.02
Boys : Attitude D	-0.08	(0.14)	0.92	-0.20	(0.32)	0.82
Younger : Attitude A	-0.18	(0.19)	0.83	-0.22	(0.22)	0.80
Younger : Attitude B	0.12	(0.12)	1.13	0.16	(0.18)	1.17
Younger : Attitude C	0.11	(0.10)	1.11	0.06	(0.16)	1.06
Younger : Attitude D	-0.04	(0.13)	0.96	0.00	(0.29)	1.00
Dutch : Attitude A	-0.07	(0.21)	0.94	-0.07	(0.24)	0.93
Dutch : Attitude B	0.21	(0.13)	1.24	0.19	(0.20)	1.21
Dutch : Attitude C	-0.05	(0.11)	0.95	-0.01	(0.16)	0.99
Dutch : Attitude D	-0.09	(0.14)	0.91	-0.11	(0.30)	0.90

Note: The coefficients are constrained to sum-to-zero. The italicized coefficients are redundant, but added for clarity. * significant at $p < .10$. ** significant at $p < .05$. *** significant at $p < .01$. Standard errors in parentheses.

The 126 students who provided an adequate explanation are again more likely to score level B or C than A or D. We find two significant relations with attitudes. Students in general secondary school are more likely to score level B, and boys are more likely to score level A. Both effects were not present for the full sample.

Conclusion

In this study, we used rubrics to assess students' citizenship attitudes. The theoretical framework describes the rationale for the three rubrics we developed, each describing four levels of attitudes towards three social tasks. The levels from A to D reflect an increasing complexity in students' attitudes towards the social tasks. In order to evaluate the use of rubrics to assess citizenship attitudes we aim to investigate (1) what level students assess their citizenship attitudes to be, (2) to what extent they are able to adequately explain their citizenship attitudes, and (3) to what extent their citizenship attitudes relate to students' individual background.

Students assessed their own attitudes by choosing the level they felt best described them and explained their choice. Figure 4 shows the odds ratios for the levels based on students' self-assessment. Figure 5 shows the results for only those students who also provided an adequate explanation. The distributions over the levels for the three social tasks show slightly different patterns. However, in all three cases students were more likely to select level B or C than A or D.

Students were most likely to score level B on attitudes towards acting democratically, and least likely to score A. This creates a picture of students generally being implicitly supportive of democratic principles and acting democratically. Boys were more likely to assess their own attitudes level A, which means they tend to more often consider democratic decision-making unproblematic and instrumental, but this relation disappeared when only considering students who provided an adequate explanation. When only considering students who provided adequate explanations students are more likely to score level B or C, but less likely to score level D. These students show (implicit) support for democratic principles and (in the case of level C) an intention to actively apply these in discussions.

Students were also most likely to score level B on attitudes towards acting socially responsible, but less likely to score level D. This creates a picture of students generally supporting the value of collaboration and a willingness to help others. They are least likely to also strive to promote these attitudes among others. When only considering students who provided an adequate explanation students remain most likely to score level B, but less likely to score level C or D. Students of whom both parents were born in the Netherlands were less likely to choose level D, but this effect disappeared when only considering students who provided an adequate explanation. Boys were less likely than girls to choose level C. When only considering students who provided an adequate explanation this difference maintained and boys were also more likely to score level A.
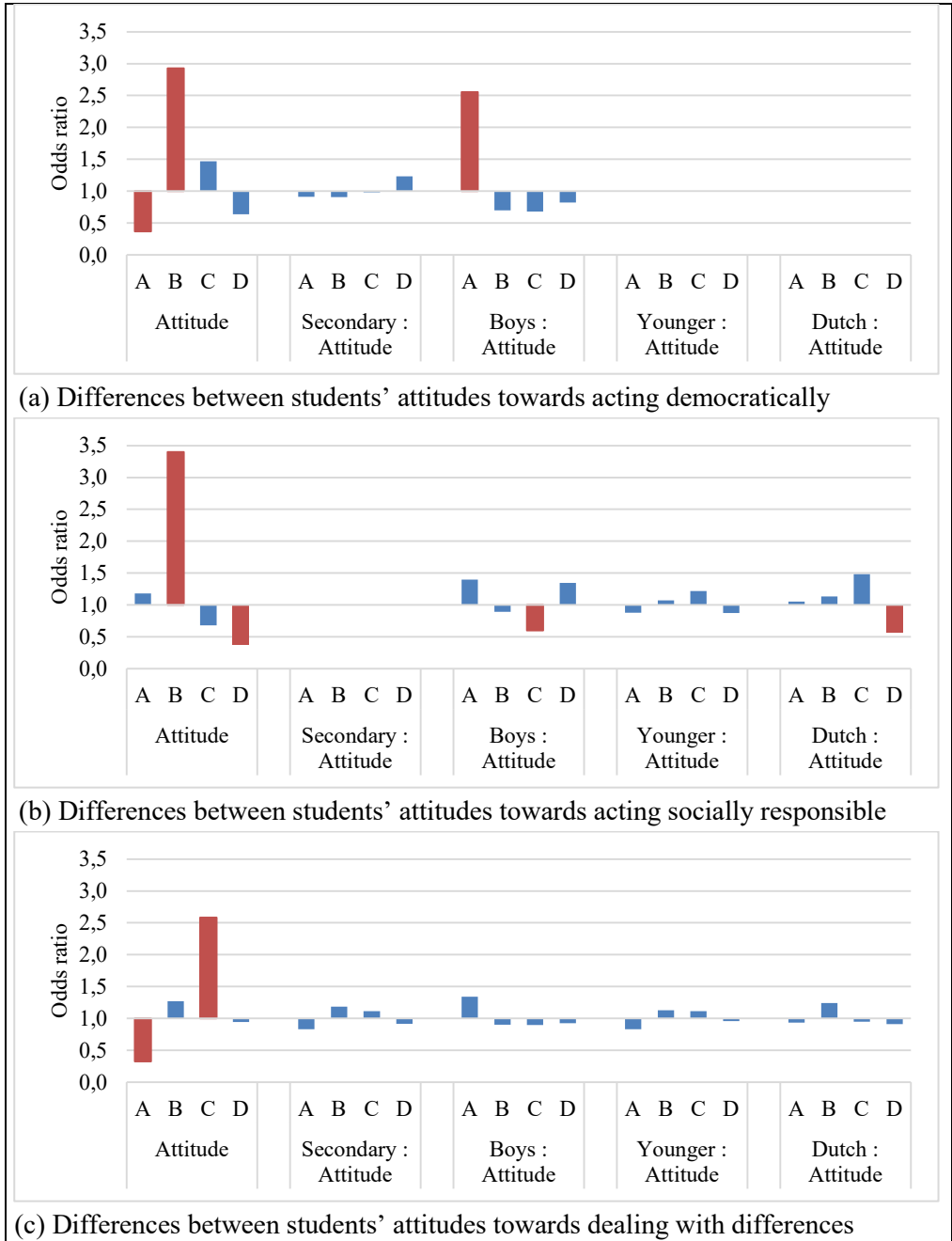
(a) Differences between students' attitudes towards acting democratically

(b) Differences between students' attitudes towards acting socially responsible

(c) Differences between students' attitudes towards dealing with differences

Figure 4. Odds ratios for students' self-assessed attitudes towards each social task.

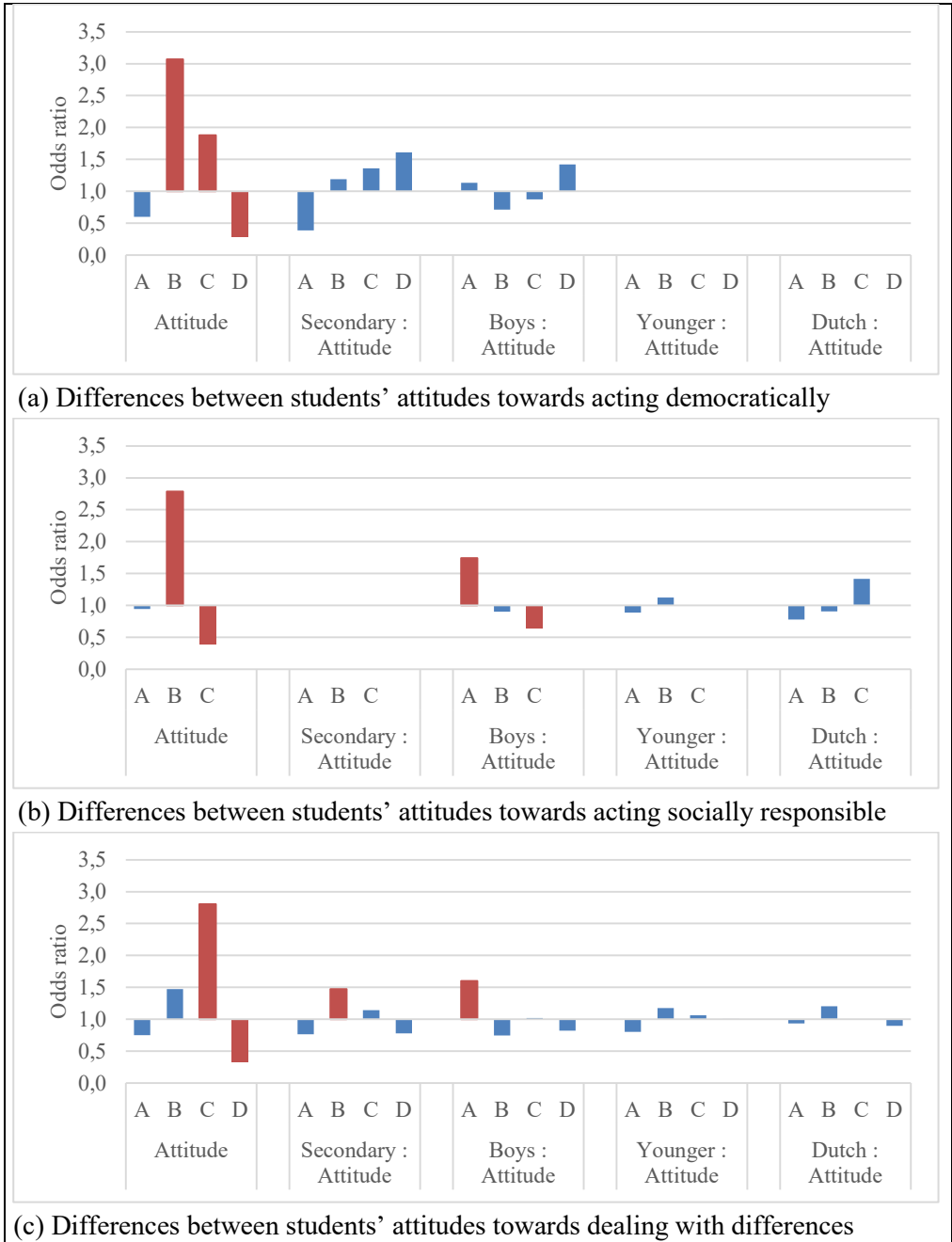Note: red bars indicate odds-ratio significantly different from 1:1 ($p < .05$).

(a) Differences between students' attitudes towards acting democratically

(b) Differences between students' attitudes towards acting socially responsible

(c) Differences between students' attitudes towards dealing with differences

Figure 5. Odds ratios for students' adequately explained attitudes towards each social task.
Note: red bars indicate odds-ratio significantly different from 1:1 (p < .05).

Students were most likely to score level C on attitudes towards dealing with differences, and least likely to score A. This creates a picture of students generally valuing and being sensitive to differences. We found no differences between students' self-assessed attitudes based on their background. When only considering students who provided an adequate explanation students remain most likely to score level C, but least likely to score level D. For this subset students in secondary school were more likely to score level B, and boys more likely than girls to score level A.

The explanations which students provided to answer why they felt they are at a certain level were adequate in nearly half the cases. In all three cases students who selected a higher level had more difficulty to provide an adequate explanation, which also reflected the higher complexity and sensitivity required for higher levels. With slightly more than half of students not providing an adequate explanation these results call for further support for students in elaborating their attitudes and supporting rationale.

Similar to survey studies using questionnaires, we found that the differences in student attitudes are generally in favour of girls. We found that girls assess their own attitudes higher than boys on attitudes towards acting democratically and acting socially responsible. When only considering students who provided an adequate explanation girls score higher than boys on attitudes towards acting socially responsible and dealing with differences. The latter findings align with those reported by Geijsel et al. (2012), who found no difference for attitudes towards acting democratically, and found that girls score higher on attitudes towards acting socially responsible and dealing with differences. Both instruments use the same theoretical framework, but different operationalisations. Using both instruments to study the same group of students could tell us more about these differences and similarities.

Various other studies found that students from higher educational tracks score higher on citizenship attitudes (e.g. Geijsel et al., 2012; Munniksma et al., 2017). Our sample included students from fourth year general secondary education (grade 10) and first year vocational education (grade 11). We only found a difference for students who provided an adequate explanation towards dealing with differences, with students from secondary education being more likely to score B. A possible explanation might be that the two tracks included in our sample are adjacent in a system with seven tracks in total, which means the differences were likely going to be small relative to other studies. Our understanding of the level of citizenship attitudes among students in vocational education is limited – even more so than for general secondary education – we are unable to provide a

straightforward explanation. Other studies have suggested that Dutch tertiary vocational education and general secondary education have grown increasingly similar over the past decades (Karsten, 2016). Applied to citizenship, this could mean that the dichotomy between lower and higher educational tracks is not as clear-cut as is sometimes suggested.

Nor did we find any differences for students' age. Because students' age was strongly related to their educational level, we used students' relative age within each educational level. Other studies also show small or no differences due to students' age, or a development of attitudes in both positive and negative directions depending on the topic (Geijsel et al., 2012; Keating et al., 2010). The current lack of understanding concerning the relationship between age and citizenship attitudes also underlines the need for longitudinal research.

We also investigated whether students' migration background related to their citizenship attitudes. We only found a marginally significant relationship for students whose parents were both born in the Netherlands being more likely to choose level C for acting socially responsible. We are reluctant to draw any conclusions from this finding. The diversity in our sample (parents were born in 57 different countries) might explain why there does not appear to be a clear relationship. Karssen et al. (2017) further showed that students of whom one parent was born abroad have scores on citizenship attitudes between those of students with both Dutch-native parents and students with both parents born abroad. Dutch students with a migration background originate from a range of countries and are now commonly second, and increasingly third generation migrants. Any relationship between student migration background and citizenship attitudes is likely to be complex.

The assessment of citizenship attitudes in this study specifically focused on its application in education. With most students assessing themselves at level B or C, teachers can use these results to identify stronger (level D) or weaker (level A) students. The fact that the students' self-evaluations and the arguments they provided can be used to assess the adequacy of the self-evaluation shows that the instrument is, in principle, relevant to learning process management because it can be used to gain insights both the existence of discrepancies between the selected level of attitudes and the quality of argumentation. We therefore conclude that the rubrics approach we developed yields results suitable for potentially supporting the learning process (cf. Black & William, 1998; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). The instrument appears less suitable for more generally descriptive goals. About half the students did not provide an adequate explanation for their selected level of attitudes, which complicates any inference on their actual

level. Questionnaires relying on a set of Likert-type items also appear to allow for the identification of smaller differences than the four-category scale on which our rubrics rely. However, in view of the patterns appearing in some of the results, a larger sample size could have also led to finding more significant differences.

Discussion

The framework for citizenship competences from which we developed the rubrics in this study has earlier been used to assess citizenship attitudes using Likert-type items on a questionnaire (Geijsel et al., 2012; Ten Dam et al., 2011). On these items, students typically indicate the extent to which they agree to a set of statements. Students who show more support for e.g. democratic principles score higher. In the theoretical framework we have described how higher levels of citizenship attitudes can also be considered from a perspective of – *inter alia* – more sensitivity for tensions between different principles (e.g. freedom and equality). By using a comprehensive assessment of students' citizenship attitudes these tensions are inherently part of the instrument. The theoretical framework describes the rationale for the four levels in more detail. The results also show that students had more difficulty providing an adequate explanation for higher levels of citizenship attitudes, where a lack of sensitivity or complexity in students' explanations would render it 'inadequate to support the self-assessed level'.

One of the concerns we anticipated was if students' self-assessment would sufficiently spread over the different levels. Most students assigned themselves level B or C for all three rubrics, with around half students scoring level B on attitudes towards acting democratically and acting socially responsible, and around half scoring level C on dealing with differences. The results show that the levels to which students assign themselves are suitable for identifying students with particularly high or low scores, as well as offering the majority of students a perspective of a higher level. Although the rubrics were developed in parallel, we cannot directly compare students' outcomes. Each student assessed his or her attitudes on only one of the social tasks, and while the rubrics were designed in the same manner, we cannot exclude that scoring high on attitudes towards one social task is easier than scoring high on another.

One of the benefits of using rubrics to assess citizenship attitudes is that they allow for students to provide personally relevant answers. Students could consider their personal situations and how they feel that citizenship attitudes are relevant to their lives. On the other hand, students' explanations were generally only a few sentences long, and since assessment relies on the quality of students'

explanations we feel that supporting and motivating students to elaborate their answers deserves continued attention. Half of the students provided adequate explanations, with most of them staying close to the description and examples given in the rubrics. We did not further investigate the lack of quality of the explanations and the underlying reasons, which most likely include the complexity of the task, overestimation, a lack of motivation for writing, and the low-stakes nature of the assignment. While it is impossible to differentiate between these factors based on the results, we do know that the rubric was presented in most classes as an auxiliary assignment, and students may not have been inclined to put in their best effort. If this conjecture is correct, it seems likely that the present findings actually undervalue the instrument's usefulness and that its application in a less exploratory and informal settings will show stronger results.

In conclusion, policy-makers continue to emphasize the importance of citizenship education. There is an on-going need for understanding the development of students' of citizenship attitudes over time and the role that schools can play in this respect. The assessment of citizenship attitudes and the availability of valid and reliable measuring instruments is an important condition for increasing this understanding. Several instruments are available in the form of various large-scale standardized surveys measuring citizenship attitudes for descriptive purposes. However this hardly, or not at all, applies to instruments focusing on facilitating the learning process (see Chapter 5). This study has shown a rubrics-based approach could make a contribution in this respect. Several issues remain that are in need of further attention. However, rubrics have shown to be a worthwhile instrument for a comprehensive assessment of citizenship attitudes in relation to the personal experiences of students.

CHAPTER 5: CONTEMPLATING MODES OF ASSESSING CITIZENSHIP COMPETENCES[1]

Abstract

Assessment of citizenship competences has become common practice following the statutory assignment of citizenship education to schools in many countries. Assessment can serve various goals. The suitability of various types of instruments depends on the alignment with the goals intended. In this paper we evaluate four types of instruments and their suitability to assess citizenship competences. Tests and questionnaires, portfolios, game-based assessment, and vignettes are each evaluated in terms of seven attributes relevant to instruments aiming to assess students' citizenship competences. Our results indicate no single type of instrument aligns with all attributes, and expanding the range of available instruments appears the best way forward so that educators and researchers can make a choice that fits their purpose. The analysis presented provides further insight into the strengths and weaknesses of particular assessment types.

Introduction

People are not born into actively engaged citizens. To effectively participate in society young people need the knowledge, skills and attitudes necessary to do so. Developing these citizenship competences involves acquiring knowledge of the functioning of a democratic society, the skills to interact with others and change perspectives, a democratic attitude and values such as responsibility, social engagement, equality and equity, as well as reflective ability by developing insights into social processes (Schuitema, Ten Dam, & Veugelers, 2008; Westheimer & Kahne, 2004). The political attitudes and beliefs that young people develop during adolescence (e.g. political interest and willingness to participate in society) largely determine their engagement in later life (Eckstein, Noack, & Gniewosz, 2012; Hooghe & Wilkenfeld, 2008; Quintelier & Van Deth, 2014). The importance of developing young people's citizenship competences is reflected in the statutory footing of citizenship education established in many countries (Eurydice, 2012).

Schools' success in teaching citizenship competences is monitored both through accountability to the various stakeholders (e.g. local environment, parents, government), as well as through in-school quality care. Assessment not only facilitates evaluating the quality of citizenship education in schools, but can also encourage a process of continuous quality improvement aimed at achieving an adequate alignment between the intentions of the school, the schools' educational practices, and the characteristics and personal goals of the specific student population. Assessment of citizenship competences can generally be said serve accountability, school improvement, and teaching and learning (Karsten, Visscher, Dijkstra, & Veenstra, 2010). The majority of countries (around 80%) partaking in the International Civic and Citizenship Education Study (ICCS) 2009 report some form of students learning assessment, and around two-third report evaluation of schools' provision in this area (Ainley, Schulz, & Friedman, 2013). In this paper we consider assessment as a special type of evaluation, and consider different instruments as modes of assessment (Dochy, 2001). To date, knowledge on student assessment in citizenship education is still preliminary, and no one way of assessing students' learning in citizenship education appears most suitable (Kerr, Keating, & Ireland, 2009; Richardson, 2010). The advantages and disadvantages of different modes of assessing students' citizenship competences are associated with the specific goals strived for.

Student assessment first and foremost serves to support students' learning. According to Jerome (2008) – following the American Psychological Association (1997) – assessment of citizenship education can have several beneficial effects on students' learning. First, learners who understand their current achievement are more likely able to plan ahead for further improvement. Second, the information derived from assessment helps teachers to provide appropriate feedback or adjust their teaching overall. Third, teams of teachers are able to compare results to discuss and improve teaching practice. Fourth, learners who are involved with their own assessment are likely to gain deeper insight into their own learning and the area in which they are learning.

Beside serving as a tool to support learning, assessment is also used to determine student performance (i.e. passing/failing grade). If desired, students' results can further be aggregated to evaluate the quality of schools in a specific domain, possibly in the form of an accountability measure for schools (Dijkstra et al., 2014b). Though evaluative applications generally incentivize student learning prior to assessment, they often do not aim to provide input to direct further learning.

For assessment of citizenship competences to serve both types of goals has proven no easy endeavor. Not in the least because citizenship competences remain a tentative concept: what constitutes 'good' citizenship appears to hold different meanings, as evidenced by the various types of citizenship distinguished (Knight Abowitz & Harnish, 2006; Van Gunsteren, 1998). If some forms of citizenship are at risk of not being acknowledged or undervalued, this would in turn affect the (construct) validity of assessment and deduct from the meaningfulness of assessment to the school as well as students themselves (Dochy & Moerkerke, 1997; Ekman & Amnå, 2012). Regarding assessing citizenship competences for student learning, the challenge for assessment then appears threefold: serving to support students' learning; evaluating students' performance; and providing meaningful insight into students' development of democratic citizenship. In this article we consider the demands this places on assessment of citizenship competences, the possible attributes of assessment instruments, and contemplate the strengths and limitations of different types of instruments. We formulate the following research question: *What opportunities do different modes of assessment offer to assess students' citizenship competences in a meaningful way, providing input to direct further learning, and appraisal of students' competence level?*

Theoretical framework

The distinction between assessment to direct further learning and appraisal of performance is generally considered as the use of assessment for learning (or: formative assessment) and assessment of learning (or: summative assessment) respectively (Taras, 2005; Wiliam & Black, 1996). We will first consider the implications of both these purposes of assessment, before exploring the implications of a developmental and meaningful assessment of citizenship competences. We then go into the issues concerning practicality of assessment in an educational setting.

Assessment of citizenship learning
Assessment of citizenship competences is of essential importance for schools to gain insight into the effectiveness of their efforts to promote citizenship learning, as it is primarily through assessment that we can find out whether a particular sequence of instructional activities has resulted in the intended learning outcomes (Wiliam, 2011). Summative assessment generally involves assigning students a (passing) grade or score. Two approaches can be taken to standardize students' performance.

Firstly, students' performance can be graded based on a set standard. Outcome descriptors are then used to describe the level of competence required to be deemed 'proficient'. Despite clear advantages of an absolute and external norm (Dijkstra & de la Motte, 2014) in the case of citizenship education there is as of yet no general agreement on what constitutes 'sufficient' or 'insufficient' citizenship competences. Citizenship competences has been shown to be a diverse concept, and the different interpretations cannot simply be hierarchically rank ordered. Moreover, studies (using quantitative data) have shown elements of citizenship competences constitute distinct aspects and cannot be combined to form a single container concept (Hoskins, Vilalba, & Saisana, 2012; Ten Dam, Geijsel, Reumerman, & Ledoux, 2011). Considering the (thus far) impossibility to determine a single scale for valuing citizenship competences, and the lack of normative agreement on proficiency levels, one of the challenges for assessment is to develop standards or norms on which to ground evaluation of the various aspects of citizenship. These standards or norms can, for example, be elaborated on the basis of what various stakeholders (e.g. educators, politicians, parents) think young people need to function adequately as a citizen in a democratic society (intersubjective assessment).

Alternatively, students' performance can be compared to that of their peers. This approach is taken in most citizenship education studies. The 2009 ICCS study, for example, consisted of 25 scales assessing various aspects of young people's citizenship competences (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008). Most of these were constructed through factor analysis of Likert-type items for which no performance-standard was set. The citizenship knowledge scale was based on an 80-item multiple-choice test. Based on the results, the researchers specified three proficiency levels of civic knowledge ability (Schulz, Fraillon, & Ainley, 2013). Although these levels are elaborated in terms of their content, they are essentially based on the performance of students partaking and thus constitute a relative norm.

Developments are still ongoing, but assessment of learning (AoL) has shown potent use to value students' citizenship competences. Specifically, AoL presents two attributes for assessment of citizenship competences

- Assessment allows comparison between students;
- Assessment allows comparison of students' performance to a norm.

Assessment for citizenship learning

Whereas assessment of learning is generally employed as a final assessment and seeks to value students' performance or progress, assessment for learning (AfL) seeks to promote the further acquisition of skills or knowledge (Dochy, 2001). AfL is defined as "the process of seeking and interpreting evidence for use by learners and their teachers to decide (1) where the learners are in their learning, (2) where they need to go and (3) how best to get there" (Broadfoot et al., 2002; numbers added). The first and second points are not unique to AfL, as they can be considered to equally apply to AoL (Taras, 2005). The third, however, marks the key difference with AoL because of the distinct attention for assessment as a process directly related to learning. Kerr et al. (2009) claim AfL is therefor particularly well suited for assessing citizenship competences, as citizenship education aims to equip students for current and future citizenship. Though arguably the use and applicability of feedback are part of the learning process, its implications for an assessment instrument are to at least provide input to facilitate learning. We therefor come to the following attribute of assessment for citizenship learning:

- Assessment provides input for students and teachers to direct further learning.

Meaningful assessment of citizenship competences

Amnå and his colleagues (Amnå, 2012; Amnå, Ekström, Kerr, & Stattin, 2009) consider a number of challenges for research into development of citizenship – or in their case: political socialization and civic engagement. They make a strong case for putting the focus on the mechanisms and processes through which development takes place. They propose a systematic approach to research which *inter alia* recognizes young people as active agents (i.e. as opposed to passive recipients to be influenced), explains interrelations between contexts, and takes a broad and longitudinal perspective (Amnå et al., 2009). We consider these points equally relevant to the issue of assessing citizenship competences.

(1) Although not without merits for societal application at the macro level, a pre-defined assessment of citizenship competences is likely not to capture what citizenship competences entail for all students (Olson, 2012). There are multiple answers to the question: "what is good citizenship?". The concept itself is furthermore constantly redefined by the influence of contemporary global dynamics through political, economic, social and demographic changes (Fischman & Haas, 2012; Knight Abowitz & Harnish, 2006). Students (and teachers and

schools) having different conceptions of what constitutes citizenship (Geboers, Geijsel, Admiraal, & Ten Dam, 2015; Leenders, Veugelers, & De Kat, 2008; Westheimer & Kahne, 2004), can be considered as inherent and – we would argue – even desirable within a pluralistic democracy. Citizenship not only entails the various forms of human activities that have a particular value and meaning for maintaining a democratic society. It also refers to the citizenship practices that are experienced as personally meaningful by students themselves. In this sense, meaningfulness encompasses both the societal and personal relevance of citizenship. Meaningful assessment of citizenship competences, adopting a broad conception of citizenship competences, allows to value different meaningful interpretations of citizenship.

(2) Notwithstanding differences in focus or definition of citizenship competences, all studies appear to agree adolescence is a critical period for the development of citizenship (Eckstein et al., 2012; Keating, Benton, & Kerr, 2011). However, though all studies appear to agree that adolescence is a critical period, little information remains on how these development take place. For instance, several studies found a notable 'dip' in several aspects of students' citizenship competences around the age of 15 (Geboers, Geijsel, Admiraal, Jorgensen, & Ten Dam, 2015; Keating, Kerr, Benton, Mundy, & Lopes, 2010). Questions remain concerning the implications of this finding, as well as whether they occur similarly for all students. Additional longitudinal research is necessary to further investigate developmental processes. Assessing citizenship development calls for measuring students' citizenship competences at several occasions, and could further be enriched by gauging students' own goals to generate insight into the processes by which citizenship competences develop.

(3) Development of citizenship cannot take place without context. Bronfenbrenner (1979) developed a model of human development which considers the development of a person to take shape in constant reciprocal interaction with (interacting) contexts; not limited to the immediate setting. Young people are continuously immerged in multiple contexts, and develop through their interaction with – and within – these structures (Dijkstra, De la Motte, & Eilard, 2014). It is the interaction with these contexts that constitutes young people's citizenship in everyday lives. More generally speaking, learning is situated in sociocultural practices: becoming a citizen is not just a matter of acquiring knowledge and skills, it also implies becoming a member of the community of practice (Lave & Wenger, 1991) and thus a change in personal identity, in the way one represents oneself to others and to oneself (Holland, Lachicotte, Skinner, & Cain, 1998). To acknowledge the significance of the contexts in which citizenship develops,

assessment of citizenship competences should be embedded in a meaningful context.

Taking a meaningful and developmental approach to assessment of citizenship competences appears to promise several distinct advantages. Three additional attributes can be assigned to assessment of citizenship competences based on this section:

- Assessment assumes a broad conception of citizenship, allowing students to elaborate their personal understanding and beliefs;
- Students' citizenship competences are assessed at several occasions, to allow insight into their development;
- Assessment is embedded in meaningful contexts.

Practicality

Beside the attributes posited thus far, the value of any assessment is determined by its use in educational practice (Gulikers, Biemans, & Mulder, 2009). Notwithstanding the range of attributes previously proposed, there is as of yet no framework to evaluate the quality of an assessment instrument in practice (Tillema, Leenknecht, & Segers, 2011). Moreover, the practical value of any instrument will largely depend on its suitability to be used to assess students' citizenship competences in schools' every day practice (Harlen, 2005). To this end a heuristic evaluation of the strengths and weaknesses appears most appropriate (Dierick & Dochy, 2001). This leads to the final attribute:

- Assessment allows advantageous use in practice.

Method

In the previous section seven possible attributes of assessment of citizenship competences were presented. In the remainder of this article we will review four types of instruments to consider their potential to meet these, on the basis whereof an informed choice for an assessment instrument can be made in the light of the specific goals strived for. Selected instruments are test and questionnaire, portfolio, game-based assessment and vignette. We focus on these types of instruments for the following reasons. Tests and questionnaires are the most applied type of instrument to assess citizenship competences, and have been rigorously developed and validated (Keating et al., 2010; Kerr et al., 2009; Schulz, Ainley, Fraillon, Kerr, & Losito, 2010; Ten Dam et al., 2011). Portfolio assessment has been

suggested as an alternative to standardized testing, because it provides a more open-ended approach (Jerome, 2008; Ledoux, Meijer, Van der Veen, & Breetvelt, 2013; Pike, 2007). Due to developments in the gaming industry, game based assessment provides new opportunities to the assessment of citizenship competences (Redecker, 2013). Several attempts have been made to develop educational games for social studies in schools, though mostly aimed at learning rather than assessment (e.g. Gordon & Baldwin-Philippi, 2014; Hanghøj, 2011; Lee & Robert, 2010). Finally, proponents of vignette instruments argue that these instruments better reflect real-life decision-making than traditional questionnaires because they mimic real decision tasks (Hainmueller, Hangertner, & Yamamoto, 2015; Wagenaar, Keune, & Van Weerden, 2012). These types of instruments present a functional selection rather than an attempt to be exhaustive. Instruments that can be considered part of an educational approach or method are excluded. Each instrument is described in terms of its application for assessing citizenship competences, and then hermeneutically evaluated in terms of the seven attributes.

Results

Tests & questionnaires

Tests and questionnaires are by far the most extensively developed instruments for assessing citizenship competences. These were employed in the international CivEd and ICCS studies, the British CELS, and the Dutch Cool5-18 study. In all these studies citizenship knowledge was assessed using a multiple-choice test (e.g. 'Why is it important for people to find out the policies of candidates before they vote?'), skills and attitudes were assessed using questionnaires mostly using Likert-scales (e.g. 'How good are you at… Imagining how another feels and taking this into consideration' or 'People should listen carefully to each other, even when they have different opinions. How well does this statement apply to you?').

Tests and questionnaires are exceedingly well suited for summative assessment, as evidenced by their long-established use as final examination throughout education. The studies cited above have shown comparison between students – or on an aggregate level between schools, or even countries – offers a range of possibilities to develop a frame of reference for citizenship competences. They assess students' citizenship on a range of components, e.g. civic knowledge, debating skills or attitudes toward equal right for immigrants. These don't add up to a single container of citizenship competence (Schulz et al., 2010; Ten Dam et al., 2011), but students' outcomes on each of these scales can be compared. In theory, comparison of student performance to a norm is fairly straightforward. However,

due to the absence of generally accepted norms this step has not yet developed (see Theoretical framework).

Tests and questionnaires score high on practicality as one of the evident benefits of these instruments is their efficiency. All students can be given the same test and within a short span of time all students in a class can typically be assessed on a range of citizenship attributes. For researchers multiple-choice items and Likert scales are easily scored, and some web-based applications even automatically generate a report on student, class and/or school level for school leaders and teachers to make use of (e.g. http://www.burgerschapmeten.nl). Comparison between students is straightforward as (in most cases) they all receive the same questions (i.e. standardization) and under more or less the same circumstances (i.e. objectivity). Finally, statistical analysis allows to investigate and optimize the reliability and validity of measurement. However, tests and questionnaires are not without disadvantages. Most of these relate to the validity of measurement. For any measurement instrument which relies at least in part on self-report, social desirability is an issue. The relationship between social desirability and citizenship is multi-interpretable. Ten Dam, Geijsel, Ledoux and Meijer (2013) show that in the domain of citizenship, social desirability not only is about 'judging oneself more positively'. The two concepts also have a substantive affinity to each other as in both cases social norms which call for the display of behavior desired by the environment play a role.

The matter of generating input to direct further learning is at least partly dependent upon the specific use in educational practice. Students' outcomes on tests or questionnaires can be used diagnostically to identify areas for further learning. At the same time the information generated by assessment will generally stay limited to a (relatively) high or low score on some scale(s). These results may be very informative in terms of where improvement is required, but they tell little about how to go about teaching or learning these.

For tests and questionnaires to provide a meaningful assessment has proven a complex endeavor. Closed-choice tests allow little room for students to elaborate on their personal understanding and beliefs. Open-choice tests are uncommon because they are more complex to grade. Items generally provide little context, which means assessment is problematic to embed in meaningful context and generates little information on the context as perceived. On the other hand, longitudinal assessment can be achieved fairly simply considering the efficiency of assessment. However, here too arise some difficulties in interpreting the meaning of change (Keating et al., 2010).

Portfolio assessment

Though portfolio instruments have a much shorter history in assessment of citizenship competences than test instruments, a multitude of such instruments have been developed over the past decades. Portfolio instruments can take a range of shapes and forms, including digital ones. Portfolio can be defined as "a purposeful collection of student work that exhibits the student's efforts, progress, and achievements in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student self-reflection" (Paulson, Paulson, & Meyer, 1991, p.60). Some (Jerome, 2008; Klenowski, 2002) claim this means portfolio assessment of citizenship constitutes not only an approach to assessment but requires a constructivist approach to teaching. As part of a portfolio a student might for example be asked to state a personal learning aim, based on given range of topics, and plan how to go about reaching that aim.

Portfolio assessment appears to be particularly well suited to meet the requirements stipulated for meaningful assessment of citizenship competences (Jerome, 2008; Pike, 2007). Portfolios provide students with room to elaborate upon their personal feelings and beliefs concerning citizenship, and reflection on these is implicit from the definition of portfolio assessment. As portfolios use products from multiple experiences, assessment is inherently embedded in contexts and students are directed to reflect on their acting in these contexts, thereby also asking them to consider how they perceived the context. As portfolios are explicitly aimed to gauge development, assessment will essentially consider competences over time. However, comparison between students may not be as straightforward as when testing at multiple occasions.

The practical strength of portfolio lies in the opportunities it provides to generate detailed information. Portfolio assessment allows students to demonstrate their learning thus providing a richer picture of students' perception than more close ended methods (Jacobson, Sleicher, & Maureen, 1999). It reveals students' understandings about learning, and allows students to interact with and reflect upon their work (Davies & LeMahieu, 2003). Portfolio assessment has furthermore been shown to be particularly suitable to evaluate programs that had flexible or individualized goals or outcomes, where there was no expressed need to compare students' performance to standardized norms (Huisman et al., 2003; McDonald, 2012). Considering these strengths attributed to portfolio assessment, it is striking how little research has examined their application to assess citizenship competences. Segers, Gijbels and Thurlings (2008) find that it is difficult to consider the general effectiveness of portfolios, because implementation and

integration into the learning environment are crucial in educational practice. Jacobson et al. (1999) warn portfolios reflect how students choose to represent themselves, and may thus provide a limited – and possibly biased – picture of what students have learned. Finally, the richness of information is also debit to the biggest drawback to portfolios. Portfolios are generally very time consuming to assess, demanding considerable effort from both student and assessor.

Portfolios are less effective in terms of summative assessment. Efforts can be made to counter subjectivity of evaluation, but comparison of students' performance to their classmates (or students from other schools) or to a norm relies on interpretation from the assessor and thus remains to some degree subjective (Jacobson et al., 1999). The fact that students can select their own evidence will lead to a diversity of portfolios, complicating the comparability of products and performance.

On the other hand, the richness of information attained through portfolio instruments does offer opportunity to inform future learning. Though it might be reasonable to assume the quality of input is dependent on the users (Segers et al., 2008), portfolio certainly offers valuable potential in this area.

Game-Based assessment

Whereas computer games are generally aimed at entertainment, 'serious games' combine this with an educational aim (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013). The vast majority of research focuses on the application of gaming to stimulate learning: game-based learning (Bellotti et al., 2013; Susi, Johannesson, & Backlund, 2007; Vandercruysse, Vandewaetere, & Clarebout, 2012; Wouters, Van Nimwegen, Van Oostendorp, & Van der Spek, 2013). By contrast, game-based assessment (GBA) entails the application of gaming technology primarily directed at assessment (Chin, Dukes, & Gamson, 2009; Mislevy et al., 2014). Three types of GBA can be employed in serious games (Ifenthaler, Eseryel, & Ge, 2012; Mislevy et al., 2014): 1) game scoring, which focuses on predefined achievements or obstacles overcome while playing; 2) external assessment, which concerns administering a test before, during and/or after students play the game; and 3) embedded assessment, which focuses on process by collecting in-game data such as log files. According to Ifenthaler et al. (2012) embedded assessment provides most benefits, as it provides detailed insight, assumes multiple measures, and allows feedback and adaptive gameplay. According to Mislevy et al. (2014, p.24) "Identifying and interpreting such data is one of the most exciting aspects of GBA, and one of the most interesting challenges to designers".

Several games have been developed that relate to citizenship competences. In *Community PlanIt*[2] players compete for points and influence (Gordon & Baldwin-Philippi, 2014). Players answer questions, contribute media, or solve problems according to their own views to gain points. Based on their points, players gain influence in decisions on real-life local community planning. The game thereby aims to create an engaged community. In *Global Conflicts*[3] students walk around in a 3d environment that represents various aspects of a regional conflict (Hanghøj, 2011). Different episodes of the game take students to perform an inquisitive role such as a journalist in areas such as Gaza. By talking to computer characters, students acquire information about the situation, on which they write an article or report after they finish playing.

Games allow students to experience interactions with the game-world, rather than being passive receivers of assessment (Squire, 2006). Though this does not necessarily mean the game-world poses a meaningful context; it shows great potential to experience situations not easily available or accessible in the real world (Dieterle & Clarke, 2008). So far, determining how students perceive this context remains difficult to determine. Though in-game behavior might provide indications of perception, further inquiry is necessary to gain a more definite picture. GBA will generally allow students to make choices, which potentially provides opportunities for students to base decisions on their own conceptions of good citizenship. Depending on the length of the game, development over time can also be assessed. In any case gaming allows continuous collection of data, and thus to map any changes to in-game behavior over time.

The practicality of using GBA in education mostly depends on whether they can be developed to reach their potential. On the one hand, game-based assessment of citizenship competences offers a range of possible strengths. Beside allowing simulation of authentic real-life contexts, simulations can be cost effective and specifically directed at citizenship competences (Corti, 2006; Dieterle & Clarke, 2008; Squire & Jenkins, 2003). Finally, through the collection of in-game data, assessment can be improved for future use, or even adapted to users real-time (Mislevy et al., 2014). On the other hand GBA poses substantial challenges. Most of these relate to the relatively underdeveloped field of GBA of citizenship competences. According to Redecker (2013) educational games in this area are set up as a teaching tool, and though they arguably support formative and

---

[2] https://elab.emerson.edu/projects/community-planit.
[3] https://www.seriousgames.net/portfolios/global-conflicts-game/.

even summative assessment, there is little evidence on the use of ICT to assess citizenship competences. Using GBA to assess citizenship competences may be relatively cost-effective once a game is available, but development costs of GBA can be substantial, and vary considerably; particularly because development is an iterative process (Mislevy et al., 2014). Moreover, as uptake of GBA remains limited in classroom practice, these practical issues concerning GBA remain persistent (Redecker, 2013).

The opportunities provided through assessment embedded in games are manifold (Ifenthaler et al., 2012): it does not interrupt the game; provides rich information about the learner's behavior through log-files; and focuses on the learning process by considering progress over a period of time. These properties would make it particularly suitable for summative assessment. Because all parameters in a gaming environment can potentially be monitored, it allows direct comparison of students' performance to each other and to a set standard. However, implementation of these developments into GBA is in its early stages, and progress is slow due to its complexity (Eseryel, Ifenthaler, & Ge, 2011; Redecker, 2013).

The use of GBA to generate input for further learning has been little documented. The richness of information generated through GBA could potentially prove very valuable for this purpose, considering all in-game behavior can be monitored. However, this implies assessment of citizenship competences through GBA is feasible. As this is yet to be the case, and a prerequisite for its use to inform further learning, assessment for citizenship learning still has a long way to go in this respect.

Vignettes

A vignette is a short carefully constructed description of a person, object, or situation, representing a systematic combination of characteristics (Atzmüller & Steiner, 2010). Vignette instruments use these descriptions to elicit respondents' judgement or interpretation. These can be presented in different forms, like keywords, dialog; narrative; cartoon; pictures; audio or video. Vignettes can be used in a range of ways. The method of delivery and set up will vary depending on the aims of assessment. Vignettes generally employ a factorial design. This means vignettes contain characteristics that are structurally manipulated to determine the main- and interaction effects of these variables (Atzmüller & Steiner, 2010). Vignettes tell stories simplifying real-life events to create an open-ended situation in which there is no one simple 'right' answer (Jeffries & Maeder, 2011).

Vignettes have been used to assess attitudinal aspects such as attitudes towards immigrants (Atzmüller & Steiner, 2010; Hainmueller et al., 2015; Turper,

Iyengar, Aarts, & Van Gerven, 2015). Wagenaar et al. (2012) used vignettes in interviews to assess several aspects of 12-year-olds' social and moral competences. An example of a brief vignette item from their study is: "Not far from town there's a nice forest. Rare plants and animals live there. A lot of people from the town go for a walk in the forest on Sundays. Chef Eddy has a plan. He wants to sell fries on one of the crossroads in the forest on Sundays. Do you think he should be allowed to do that? Indicate why you think so. Can you think of any counter- arguments?"

Though it will very much depend on the modus in which vignettes are used for assessment (e.g. interviews or written multiple choice), the use of vignettes provides opportunities to elicit more than a selected final answer as it can also probe students' argumentation. It thereby provides input of students' underlying reasoning and beliefs, which can provide input for future teaching and learning.

Similar to tests and questionnaires, vignettes are well adaptable to summative assessment. Assessment is structured in an attempt to provide an objective and reliable measure. Use of open-ended questions (or interviews) further allows to assess students' argumentation. However, considering vignettes generally do not posit a 'right' or 'wrong' answer the comparison to a set standard is somewhat more complicated than closed-choice methods. Similar to tests and questionnaires, vignettes do allow for comparison of results between students.

Vignettes can provide students room to elaborate on their personal understanding and beliefs concerning citizenship. However, modes that will allow such elaboration (such as interviews) are time consuming, and would suggest a trade-off between richness of information and opportunities for standardization. When more information on students' beliefs is collected, mapping development over time would appear feasible. Considering context is inherent to vignettes, they provide ample opportunity to embed assessment in a context meaningful to students, and questioning students' perception as well as posterior analysis allow to elaborate the context as perceived.

The use of vignette instruments offers several practical benefits. Vignettes assessing attitudes are less biased against social desirability in the sense of 'judging oneself more positively' and yield more exact measurement of attitudes than questionnaires (Auspurg, Hinz, & Liebig, 2009). Vignettes can contain a complex set of characteristics, reflecting more accurately the complex situations in which citizenship competences are expressed (Auspurg et al., 2009). According to Jeffries and Maeder (2011, p.162): "Vignettes are effective because they are brief and relatively easy to construct and administer, provide a useful focus and stimulus for discussion, are valuable in addressing difficult-to-explore and sensitive topics, can be used with individuals and groups online and in the classroom, and reflect real-

life contexts and problems". On the other hand vignettes do have some drawbacks. Complex descriptions of situations might not be suitable for students of all ages and levels, and to be used in interview settings requires considerable time from the assessor.

Discussion

This paper has evaluated several types of instruments on their potential attributes for assessing citizenship competences. We distinguished four categories of attributes to consider an instrument's potential: suitability for summative assessment, suitability for formative assessment, meaningfulness to students, and practical applicability. A total of seven attributes were formulated on which instruments were hermeneutically evaluated. Four types of instruments were considered: tests and questionnaires, portfolios, game-based assessment, and vignettes. Our aim has been to provide a comprehensive review rather than attempting to be exhaustive. Nor has it been our intention to evaluate the quality of assessment instruments representative for a type of assessment. We hope the insights shared in this paper help educators and researchers making an informed choice for a particular assessment instrument that fits their purpose.

Tests and questionnaires are commonly used in combination and are the most used instruments to assess citizenship competences. Their efficiency, objectivity and standardization make them highly practical, and particularly well suited for summative assessment. Although students' results can be used to adapt future teaching, the standardized nature of tests means they generate less personal information, i.e. these instruments have difficulty probing students' underlying beliefs or argumentations. Consequentially, creating a meaningful assessment can prove a laborious endeavor.

Portfolios take a wholly different approach to assessment. Students are typically instructed to collect evidence of their competence development. Portfolio assessment hereby aims for assessment to be both personally and socially meaningful. The potentially rich information generated through portfolio assessment can be used as input to direct further learning. However – considering standardization of portfolios is detrimental to its open-endedness – objective, summative assessment is complicated by its interpretative and therefor subjective evaluation. Finally, portfolio assessment is time consuming for both the assessor and the students which compromises its practicality.

Game-based assessment presents a relatively new avenue in educational assessment, and for assessing citizenship competences particularly. The complex

multi-user environments generated in today's games show opportunities to create meaningful situations in which students can take on citizenship roles not easily accessible in real life. Because all parameters – including student behavior – can be monitored, comparative summative assessment appears feasible already, and assessment of performance compared to a norm seems possible. Formative assessment appears more tentative, as it remains unclear whether in-game behavior can be transferred to real life situations. The largest drawback however, is that for game-based assessment of citizenship competences to realize its potential still requires considerable development.

Vignette instruments attempt to replicate the complexity of real life through detailed descriptions of situations. Students' responses to these situations allow insight into their citizenship beliefs and argumentations. These can be used to provide input for future learning and teaching. The standardization of vignettes also allows to compare pupils to each other, and to a lesser extend to a set standard. For vignettes to provide a more meaningful assessment a trade-off with standardization and efficiency appears to emerge. Vignettes used in interviews provide rich information but are time consuming, and when vignettes get very detailed only few can be used because of the demand they place on student and assessor. Finally, practicality is limited as vignettes are complex to develop and rarely designed for classroom assessment.

This review shows none of selected assessment approaches has all attributes desired for assessment of citizenship competences to serve the specified purposes. Each mode of assessment has its strengths and weaknesses. Though reviewing the different types of assessment separately, we do not believe that these approaches ought to be considered mutually exclusive. In social science research, triangulation is generally considered an effective way of validating research findings through different approaches. Similarly, assessment of citizenship competences could benefit from using multiple methods of inquiry whether or not combined in one and the same assessment. Our analysis highlights several opportunities for complementary approaches. For instance, in addition to a portfolio type assessment, students could be asked to complete a test to assess their knowledge of a certain topic while also being asked to reflect on a related personal experience. Similarly, vignette items could be added to tests and questionnaires to include more real-life inquiry.

Considering instruments currently developed consist mostly of tests and questionnaires, the development of portfolio assessment could augment insights into the development of citizenship by attuning to a more meaningful type of assessment. Further development of game-based assessment and vignettes also

promises several distinct advantages, and can increase the range of instruments from which educators and researchers can take their pick. By extension, expanding the scope of instruments and modes of assessment, also beyond the types of instruments discussed here, appears both desirable and necessary for improving assessment of students' citizenship competences as well improving educational practice.

This paper focused on the assessment of student citizenship competences, particularly in educational settings. From the responsibility assigned to schools to teach citizenship education it follows that schools ought to be evaluated in terms of their success in promoting citizenship competences. After all, schools do not operate in a vacuum and the quality of citizenship education is a matter of public interest. Dijkstra et al. (2014b) consider three models for educational inspectorates to assess school effectiveness in the social domain: focusing on school improvement, the process of teaching and learning, and student outcomes. All three these models require insights into students' competences, focusing on learning or results. In many countries schools are therefore mandated to assess students' competence level (Ainley et al., 2013). Schools would do well to consider what approach to assessment best aligns with their teaching approach as well as external accountability obligations. Assessment of citizenship competences could benefit from a deliberative school approach, where strengths and weaknesses are taken into account and testing is augmented with other types of assessment.

CHAPTER 6: CONCLUSION AND DISCUSSION

This dissertation centers on the assessment of citizenship competences. In the four studies we conducted assessment of citizenship is approached from different angles. This final chapter presents an overview of our findings. We will first briefly review the main findings per study from the previous chapters. We then turn to the contributions these studies make to our understanding of assessing citizenship competences, as well as discussing their limitations. We conclude by presenting the implications of our findings for future assessment of citizenship competences.

Main findings

Results on assessment of citizenship competences reflect the education policy context

Student assessment results can be aggregated to class, school or even national level. Because the International Civics and Citizenship Education Study (ICCS) uses a nationally representative sample, it allows results to be compared between countries. This allows us to evaluate if students' results on such a survey are reflective of national policy and discourse on citizenship education. The first study of this dissertation considers the aims and outcomes of citizenship education policy in Norway, the Netherlands, Scotland and Sweden. While these countries share many similarities, they each have a distinct approach to the issue of citizenship education. Through analysis of policy documents, research reports, inspection reports and scientific papers for each of these four countries a distinct picture of the state of citizenship education is drawn. These images are then compared to students' results on five selected scales from ICCS 2009: civic knowledge and understanding, support for democratic values, support for equal rights for all ethnic/racial groups, expectations to participate in elections, and reports of civic participation at school. In all four countries citizenship education received considerable attention over the past years. However, there are notable differences in the degrees to which certain issues are discussed. Because ICCS data is cross-sectional the aim is not to drawn causal inferences, but rather to investigate associations. Because Scotland did not partake in ICCS 2009, Scottish students' outcomes could not be compared.

  The attention for citizenship education in the Netherlands appears strongly motivated by notions of social integration of immigrants. Citizenship education policy in the Netherlands is strongly directed at 'living together'. However, various studies indicate educational provision shows little development towards these aims.

Students in the Netherlands scored similar to the international average on citizenship knowledge, and below average on all other four selected aspects of citizenship competences. The results also show differences between schools are much larger in the Netherlands than in Norway or Sweden; likely due to the Dutch tracked school system.

Citizenship education in Norway appears not to receive as much attention from a policy perspective as in the other three countries. Outcomes on assessment of students' citizenship competences have historically been high, and education policy appears mostly to focus on cognitive outcomes. Mechanisms to promote students' citizenship competences can however still be identified embedded in the curriculum, for example in subjects such as social studies. Students in Norway scored higher than average on all five measures of citizenship competences.

Citizenship education policy in Scotland appears to focus strongly on student participation. Development of citizenship is one of five central priorities of education, and education for citizenship is promoted cross-curricular and as part of social studies. Based on these arrangements Scottish students can be expected to score average to high on most citizenship competences outcomes. Scotland did not partake in the ICCS, which can be considered to reflect a focus on the process of citizenship education rather than its outcomes (cf. Dijkstra & de la Motte, 2014).

Citizenship education policy in Sweden can be considered to stem from a longstanding tradition of promoting social values, though in recent decades policy has grown more liberal. The prominence of these values inside and outside education is still recognized. Swedish students score in between Dutch and Norwegian students, scoring higher than average on citizenship knowledge and support for equal rights for ethnic/racial groups, average on support for democratic values and reports of civic participation at school, and below average on expectations to participate in elections.

The aggregation of students' results on the ICCS tests and questionnaires allow comparisons between countries (Schulz et al., 2010). The analyses suggest that students' outcomes on citizenship competences reflect the discourse on (the effectiveness of) citizenship education policies. The study shows countries take distinctly different approaches to citizenship education policy, and that these are relevant to take into account when interpreting (aggregated) measures of students' citizenship competences (cf. Hahn, 1999). Put more simply: citizenship education policy matters.

Rubrics can be used to assess citizenship competences

Most studies of students' citizenship competences use a survey design and employ (typically multiple-choice) tests and questionnaires. There have been frequent calls to expand the range of available instruments to other assessment approaches (Jerome, 2008; Kerr, Keating & Ireland, 2009). The second study focuses on an alternative approach to assessing citizenship competences. To this end we developed a new assessment instrument for citizenship competences. Because the instruments available to date typically employ multiple-choice tests and questionnaires to assess citizenship competences, we developed an instrument that could potentially offer room to students' own interpretation and support their further learning: rubrics (cf. Andrade, 2005; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). We developed three rubrics on three 'social tasks': acting democratically, acting socially responsible, and dealing with differences; i.e. exemplary aspects of citizenship relevant to young people's daily lives (Ten Dam et al., 2011). These social tasks can also be considered to reflect the focus in Dutch citizenship education on 'living together' (Ten Dam & Volman, 2007), which is where we conducted our study. Each rubric consisted of three dimensions: knowledge, attitudes, and skills. All three dimensions were described at four levels: A through D. The full rubrics are included in Appendix A; the following is a (very) brief description of how the different levels can be characterized.[1] The levels from A through D have to be considered as a categorical scale. Each consecutive level presupposes the lower levels have also been achieved.

Knowledge level A describes knowledge of fundamental principles, such as to do no harm to others. Level B describes understanding the consequences of these principles, such as understanding the limits to free speech. Level C describes application of knowledge, such as the implications of living in a welfare state. Level D describes evaluation of knowledge, such as evaluating the causes and effects of political decisions.

Attitudes level A describes attitudes that can be considered legitimate, but assumes living together in society – the focus of the investigated attitudes – as unproblematic. Level B reflects what can be considered attitudes that are implicitly supportive of democratic principles. Level C describes attitudes that can be considered explicitly and deliberately supportive of democratic principles. Level D

---

[1] The rationale behind the levels of knowledge, attitudes, and skills are described in more detail in Chapters 2 and 3 of this dissertation.

describes attitudes that are sensitive to complex situations, and supportive of promoting democratic attitudes among others.

Skills level A describes the ability to shape one's own behavior, such as sticking to your own opinion. Level B describes the ability to shape behavior in relation to others, such as having a discussion. Level C describes the ability to adjust behavior in relation to others and circumstances, such as adjusting to cultural differences. Level D describes the ability to act purposely in demanding situations, such as when attempting to convince others.

The rubrics we designed are intended to assess students' citizenship competences in educational settings. The development of the contents of each rubric has therefore been an iterative process. We departed from a consideration of what conceptual differences might define different levels based on theory as well as empirical results. The resulting draft versions of these rubrics were discussed with teachers, students, and researchers and adapted based on their feedback. Particularly the examples presented in the rubrics were often adjusted based on these meetings.

Because our pilots showed younger students had trouble understanding the rubrics, we assessed students who were generally 15 years or older. Students in the Netherlands in fourth year general secondary education (grade 10) and first year tertiary vocational education (grade 11) took part in the study. No earlier studies have been conducted of the citizenship competences of these students, although 'civics' and 'career and citizenship' have been part of all Dutch general secondary and tertiary vocational education tracks respectively for some time. We tested the rubrics using two approaches: 115 students completed a reflection assignment based on which their knowledge and attitudes were assessed using one of the rubrics; 601 students assessed their own citizenship knowledge, attitudes and skills based on one of the rubrics.

The results for using rubrics to assess students' reflection reports showed almost all students could be assigned to a level based on simple-majority agreement between three assessors. Inter-rater agreement was moderate to substantial. Most students scored A or B on both knowledge and attitudes, around 10 to 15 percent scored C, and none scored D. This distribution was similar to the self-assessment when only considering students who provided an adequate explanation. The rubrics appeared suitable for assessing student knowledge and attitude, but did show a dependency on the quality of students' explanations to their answers.

The self-assessment approach shows almost all students are able to choose a level that they feel best describes their level of knowledge, attitudes, or skills.

Students were overall most likely to choose level B or C. Over 90 percent of students also provided a relevant explanation with their choice. When only considering students who provided adequate explanations, they were most likely to choose level B or A. However, only one in four explanations for knowledge and skills was considered adequate; for attitudes this applied to half of explanations. Because it is impossible to infer if inadequate explanations are due to overestimation, lack of motivation, lack of writing skills, or other reasons self-assessment of citizenship competences using rubrics can be considered more suitable for citizenship attitudes than for knowledge and skills, and students need support in self-assessing – and particularly explaining – their competences.

Rubrics provide insight into students' citizenship attitudes
Students' attitudes take a central place in the development of citizenship competences. They affect the collective by helping to sustain democratic societies by supporting democratic ideas, values and practices (cf. Dahl, 2000; Eckstein, 1966; Welzel, 2007) and they affect the individual's ability to act effectively as socially competent citizens in a democratic society (Ten Dam & Volman, 2007).

Students' citizenship attitudes are typically reported based on their agreement to a set of Likert type items. The results of the second study suggest rubrics may be particularly suitable for the assessment of citizenship attitudes. In the third study we therefore look more closely at the self-assessment of citizenship attitudes using rubrics, and more specifically at what assessment of citizenship attitudes using rubrics teaches us about students' attitudes towards acting democratically, acting socially responsible, and dealing with differences. The self-assessed level of attitudes and explanations of 567 students are analyzed, as well as differences between students based on their background characteristics. Students were 16 years old on average, and in fourth year general secondary or first year tertiary vocational education.

Students were most likely to score level B on attitudes towards acting democratically, and least likely to score A. Students appear generally implicitly supportive of democratic principles and acting democratically. When only considering students who provided adequate explanations students are more likely to score level B or C, but less likely to score level D. These students show (implicit) support for democratic principles and (in the case of level C) an intention to actively apply these principles in discussions.

Students were also most likely to score level B on attitudes towards acting socially responsible, and least likely to score level D. Students appear generally supportive of the value of collaboration and willing to help others. They are least

likely to also strive to promote these attitudes among others. When only considering students who provided an adequate explanation students remain most likely to score level B, but less likely to score level C or D.

Students were most likely to score level C on attitudes towards dealing with differences, and least likely to score A. Students appear to generally value being sensitive to differences. When only considering students who provided an adequate explanation students remain most likely to score level C, but least likely to score level D, which means they are least likely to also strive to promote these attitudes among others.

The results differ somewhat between the three social tasks, but overall students are most likely to assess their attitudes at level B or C. This remains the case when we only consider students who provided an adequate explanation (though the percentages of students at higher levels does decrease slightly).

The evaluation of using rubrics to assess students' citizenship attitudes focused specifically on their application in education. The fact that the students' self-evaluations and the arguments they provided can be used to assess the adequacy of the self-evaluation shows that the instrument is, in principle, relevant to learning process management because it can be used to gain comprehensive insights into students' citizenship attitudes, as well as to identify the existence of discrepancies between the selected level of attitudes and evaluate the quality of students' argumentation (cf. Black & William, 1998; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). In relation to the latter point made, the overall quality of explanations left something to be desired with only half of students providing an adequate explanation.

Using rubrics to assess citizenship attitudes provides a conceptually different measure of citizenship attitudes than a set of Likert type items. Higher levels of attitudes are conceptualized as increasing complexities of considerations and sensitivities. The results of this study show students are generally supportive of democratic principles, value helping others, and value showing sensitivity to others, but most do not yet consider their attitudes sensitive to exceptions or complexities. We found only some small differences between students based on their age, gender, education level, or migration background.

Different approaches to assess citizenship competences are suitable
Assessing citizenship competences can serve different purposes. While these can be categorized in many different ways, a common distinction is that between assessment *of* learning and assessment *for* learning (Broadfoot et al., 2002; William & Black, 1996). Essentially, both these types of assessment aim to provide some

form of judgement, where assessment for learning also aims to provide input to support further learning (Taras, 2005). To further develop opportunities for both assessment *of* and *for* citizenship learning it is relevant to consider the extent to which different instruments can facilitate these aims.

In the fourth and final study we evaluate four approaches to assessing citizenship competences: tests and questionnaires, portfolios, game-based assessment, and vignettes. Seven criteria for evaluating assessment instruments are formulated, based on four central concepts. Based on assessment *of* citizenship learning, assessment should (1) allow comparison between students, and (2) allow comparison of students' performance to a norm. Based on assessment *for* citizenship learning, assessment should (3) provide input for students and teachers to direct further learning. Based on creating a meaningful assessment of citizenship competences, assessment should (4) assume a broad conception of citizenship, allowing students to elaborate their personal understanding and beliefs, (5) assess citizenship competences at several occasions, to allow insight into students' development, and (6) be embedded in meaningful contexts, in which students consider citizenship competences relevant to their lives. Finally, based on practicality, assessment should (7) allow advantageous use in educational practice.

Tests and questionnaires are commonly used in combination and are the most used instruments to assess citizenship competences. Their efficiency, objectivity and standardization make them highly practical, and particularly well suited for summative assessment. Although students' results can be used to adapt future teaching, the standardized nature of tests means they generate less personal information, i.e. these instruments have difficulty probing students' underlying beliefs or argumentations. Consequentially, creating a meaningful assessment can prove a laborious endeavor.

Portfolios take a wholly different approach to assessment. Students are typically instructed to collect evidence of their competence development. Portfolio assessment hereby aims for assessment to be both personally and socially meaningful. The potentially rich information generated through portfolio assessment can be used as input to direct further learning. However – considering standardization of portfolios is detrimental to its open-endedness – objective, summative assessment is complicated by its interpretative and therefor subjective evaluation. Finally, portfolio assessment is time consuming for both the assessor and the students which compromises its practicality.

Game-based assessment presents a relatively new avenue in educational assessment, and for assessing citizenship competences particularly. The complex multi-user environments generated in today's games show opportunities to create

meaningful situations in which students can take on citizenship roles not easily accessible in real life. Because all parameters – including student behavior – can be monitored, comparative summative assessment appears feasible already, and assessment of performance compared to a norm seems possible. Formative assessment appears more tentative, as it remains unclear whether in-game behavior can be transferred to real life situations. The largest drawback however, is that for game-based assessment of citizenship competences to realize its potential still requires considerable development.

Vignette instruments attempt to replicate the complexity of real life through detailed descriptions of situations. Students' responses to these situations allow insight into their citizenship beliefs and argumentations. These can be used to provide input for future learning and teaching. The standardization of vignettes also allows to compare pupils to each other, and to a lesser extend to a set standard. For vignettes to provide a more meaningful assessment a trade-off with standardization and efficiency appears to emerge. Vignettes used in interviews provide rich information but are time consuming, and when vignettes get very detailed only few can be used because of the demand they place on student and assessor. Finally, practicality is limited as vignettes are complex to develop and rarely designed for classroom assessment.

All four assessment approaches have their strengths and weaknesses in relation to these seven criteria. Tests and questionnaires show strong suitability for assessment of learning, but can have difficulty creating a meaningful assessment. Portfolios provide a rich assessment that can facilitate meaningful assessment, but often struggle with reliability and time constraints. Game-based assessment appears promising to create meaningful summative assessment, but still requires considerable development to deliver on these promises. Vignettes show a trade-off between opportunities to reconstruct meaningful real life situations and high reliability summative assessment, but they are meticulous to develop.

Contributions

Taken together the studies presented here provide three contributions to the development of citizenship competences in education and their assessment. In this section we consider the range of functions that assessing citizenship competences can fulfil, the significance of our newly developed approach to assessing citizenship competences, and the value of defining levels of citizenship competences in rubrics.

Functions of assessment

In this dissertation we discuss two purposes of assessment that are commonly considered: assessment of learning aims typically aims to measure student achievement compared to their peers or a set standard; assessment for learning aims to provide input to support further learning. These purposes provide a broad framework from which assessment instruments can be considered and evaluated to what extent they facilitate a certain purpose. However, the specific function an assessment fulfills can often be considered more specifically when looking at the assessment approach by which an instrument is used. We consider these functions at three levels: that of individual students, a group of students at class or school level, and across schools at the education system level.

Assessment of individual students' citizenship competences can fulfil a range of functions. Assessment might for example serve to generate a report card grade, reflecting students' level of proficiency or effort. These might help both students and teachers to monitor progress over time, and plan their actions accordingly (cf. Jerome, 2010). Assessment might also be used to motivate students to study, or even to convey that a certain topic is important because 'it will be on the test' (cf. Richardson, 2010). If the assessment takes place prior or during learning, the outcome might also be used to inform students to set their learning goals.

Teachers might also use assessment results to evaluate their teaching approach, or compare teaching practices among each other or over time. When using standardized assessment, results might be compared between schools to evaluate school practice and adjust these accordingly. Other stakeholders might also be interested in reviewing these results as an indication of school quality, such as school leaders, the school board, parents, or the inspectorate of education. These results might be used to monitor student or school progress, to base decision making on, for evaluation, or school accountability (cf. Dijkstra & de la Motte, 2014). The first study of this dissertation shows, in line with others (cf. Schulz et al., 2017), that differences between schools are strongest for citizenship knowledge, and differences between schools differ between countries.

Assessment results may also be used to evaluate students' citizenship competences at the school system level (cf. Schulz et al., 2017). Assessment results may be used to introduce new policy initiatives, or to evaluate effectiveness of existing policy. Results may also be used for school accountability or to identify best practices or failing schools (cf. Dijkstra & de la Motte, 2014). The first study of this dissertation shows students' results can be considered to reflect education policy and discourse on citizenship education.

Assessment instruments are typically designed for a specific purpose. However, especially considering the scarcity of instruments available for the assessment of citizenship competences, results are often used for multiple purposes. This may result in discussion on the suitability of assessment instruments for certain purposes (cf. Campbell, 1979). For example, based on students' citizenship knowledge on ICCS 2009 and 2016 the Dutch Ministry of Education recently concluded the objectives for citizenship education had been achieved, but citizenship education could be further improved (Ministerie van Onderwijs, Cultuur & Wetenschap, 2017). At the same time, the minister of education called the results of the ICCS study "useless" (Tweede Kamer, 2018). However, we would not suggest assessment instruments cannot fulfil multiple functions. For example, a teacher testing if students understood an explanation of what constitutes a 'welfare state', might also want to use students' results to evaluate if his explanation was clear. Rather, the solution would be a more careful consideration of the suitability of assessment for specific purposes. The fourth study of this dissertation evaluates to strengths and weaknesses of using tests and questionnaires, portfolios, game-based assessment and vignettes to assess citizenship competences, and shows certain types of instruments may be more or less suitable for particular purposes. One of the ways forward appears to expand the range of instruments available to choose from (cf. Kerr et al., 2009).

New approach to assessing citizenship competences
The second and third study center on a central contribution this dissertation has endeavored to make: a new approach to the assessment of citizenship competences. The development of rubrics to assess citizenship competences took place over the course of two years, during which teachers, students and researchers were interviewed and asked to reflect on the contents of each rubric. Instruments available for the assessment of citizenship competences to date show strong similarities (cf. Keating et al., 2010; Schulz et al., 2010; Ten Dam et al., 2011). The instruments are generally aimed at the assessment of learning (i.e. summative assessment). A new type of assessment instruments could produce new insights into the development of citizenship competences of young people. Moreover, the instruments currently available are typically designed for scientific studies of students' citizenship competences. Instruments that facilitate the provision of feedback to students and teachers have the potential to support the future acquisition of competences (Black & William, 1998).

The second study of this dissertation investigates two applications of the rubrics we developed. The first approach asked students to self-assess their

competences based on the rubrics we presented them. The results show almost all students are able to choose a level they feel best reflects their proficiency, and the majority of students provided relevant explanations. However, the number of students who provided adequate explanations to support the level they selected was much lower, particularly for knowledge and skills only one in four explanations could be considered to adequately the level that students selected. The second approach asked students to complete a reflection assignment. Based on their answers students' citizenship knowledge and attitudes were assessed using one of the rubrics. Assessors agreed moderately to substantially on the levels of knowledge and attitudes that students were assigned.

Because the second study showed students were most successful in providing an adequate explanation to their citizenship attitudes, the third study of this dissertation looked at the self-assessment of citizenship attitudes using rubrics in more details. The results show students are generally implicitly supportive of democratic principles, and explicitly supportive of differences between people. Granted this is a first attempt, this application could allow students, teachers and researchers a more comprehensive insight into students' citizenship attitudes, which could also help students and teachers to inform further learning. Rubrics thus appear suitable for the assessment of citizenship competences, and for assessment of citizenship attitudes in particular.

Defining levels of citizenship competences

When assessing citizenship competences a higher score is generally perceived as 'better'. This also shows in the first study of this dissertation, where higher aggregated scores of students are often perceived as less problematic in discussions on citizenship education policy. This rationale appears plausible when applied to citizenship knowledge, where a deeper understanding of democratic principles and the role of institutions will generally help citizens better understand democratic processes. But is having more trust in political institutions always a good thing? And what does it mean when students say they are very good (as opposed to not very good) at adopting a socially just position? For these and other aspects of citizenship competences higher scores by students are also generally considered 'better'. By putting into words different levels of citizenship competences in rubrics this instrument contributes to considering in more detail what different levels of competences entail, and how these competences might develop.

Earlier studies have endeavored to identify different levels of citizenship knowledge, which greatly helped us when conceiving our levels of citizenship knowledge. Based on results from the ICCS 2009 and 2016 civic knowledge tests,

Schulz and colleagues defined different levels of citizenship knowledge (Schulz et al., 2010, 2013, 2017). Wagenaar et al. (2011) asked experts to rate a set of citizenship knowledge items on difficulty for students with minimal, sufficient and advanced proficiency. We also looked at what items on the Citizenship Competences Questionnaire students typically score lower on (Geboers et al., 2015; Geijsel et al., 2012; Ten Dam et al., 2011). We further used Bloom's taxonomy to consider different levels of knowledge (Bloom, 1956).

Similar to citizenship knowledge, we conceptualized higher levels of attitudes and skills by considering what constituted a more complex proficiency, e.g. being sensitive to people's differences in situations where these might lead to tensions, or also being able to reach consensus in a heated debate. We also looked at what items on the Citizenship Competences Questionnaire students typically score lower or higher on (Geboers et al., 2015; Geijsel et al., 2012; Ten Dam et al., 2011). The contents of the distinct levels of citizenship competences are described in more detail in Chapters 3 and 4 of this dissertation, and included in Appendix A. These levels of citizenship competences are more likely to mirror what would constitute more proficient citizenship competences in daily practice, since they also reflect the different complexities in which social tasks can present themselves and the competences necessary to effectively perform these (cf. Ten Dam & Volman, 2004).

Regarding higher levels of citizenship competences as increasing complexity also presents opportunities to more closely reflect young people's development of these competences. Earlier studies have shown students' support for citizenship attitudes or feelings of self-efficacy do not develop linearly over time (Geboers et al., 2015; Keating et al., 2010). Considering and putting into words what increasing levels of proficiency might look like could help both teachers and students shape learning experience towards developing these competences.

Limitations

Of course the studies presented here are not without limitations. We discuss two of these in more detail.

Single occasion measurements

The four levels defined in the rubrics might suggest some threshold level that students attain on each component, on their way to level D. However, while the levels were designed to help inform further learning, they were not conceived as a

route map for students to systematically progress through. The second and third study of this dissertation investigate the implications of using rubrics to assess citizenship competences. The data collected and analyzed for these studies consists of single occasion measurements. The use of rubrics to support students' learning of citizenship competences needs further – longitudinal – study. This call for longitudinal study of students' citizenship competences is often voiced, and such studies are scarce (see for example Amnå, 2012; Geboers et al., 2015; Keating et al., 2010; Nieuwelink, 2016).

Role of language ability
Earlier studies have shown a relation between students' language ability and their development of citizenship competences (Eidhof et al., 2017). Considering the social nature of citizenship, it is not surprising these two concepts are considered strongly related (Starkey, 2005). Citizenship essentially concerns being part of a community, and performing social tasks (i.e. acting democratically, acting socially responsible and dealing with differences) requires the language ability to do so. Language ability can also be considered to play a role in the assessment of citizenship competences; particularly reading ability and vocabulary can be considered to play a role in understanding the items that students have to answer in tests and questionnaires. These issues are not unique to the assessment of citizenship competences, and can also be considered to affect other assessment instruments relying either on students' reading or writing.

Reading ability also appeared to play a role in students' comprehension of the rubrics we developed. During the pilot particularly teachers indicated that students would have difficulty carefully reading and understanding the full contents of the rubrics. This was one of the reasons why we chose to iteratively develop the rubrics, and why we chose to assess students who were somewhat older than in other studies of students' citizenship competences. Because around 90 percent of students provided explanations that could be considered relevant to the social task and dimension (i.e. knowledge, attitudes or skills) we would conclude students in our study understood at least the essentials of the rubrics. The role of reading ability both in the development and assessment of citizenship competences still needs further investigation, but would appear not to have played a larger part for our rubrics than for other assessment of citizenship competences.

Both the self-assessment and the reflection assignment relied on student writing, and we often would have liked students to elaborate more clearly on their answers. Students' writing ability may interfere with assessment of citizenship competences when students are unable to elaborate their views. When students

were asked to explain why they selected a certain level they felt best described their proficiency, they often copied parts of the rubric rather than providing their own account. Students' answers to the reflection assignment varied in length, and one of the challenges remains to get students to elaborate their answers. The cognitive load of the exercises may have played a role here. However, the tasks were presented as low-stakes to the students, which may have exaggerated these issues.

These limitations could have led to an underestimation of students' citizenship competences. For the application of rubrics to assess citizenship competences it means student should particularly be supported in giving their explanations. Additional research could investigate if assessment approaches less reliant on students' writing are feasible.

Implications

Take citizenship education context into account

Citizenship education has been on the policy agenda of many countries for the past decades. The first study of this dissertation shows the policy problem is often discussed in the national context, and these show both similarities and differences between countries. These are relevant to take into account because they not only direct policy development of educational practices, but also reflect students' citizenship competences (Hahn, 1999). As Westheimer and Kahne (2004) discuss, citizenship education can have different aims in terms of the kind of citizenship that is promoted. Leenders, Veugelers and de Kat (2008) shows these different aims can also be identified among teachers. These issues are relevant to take into account when assessing citizenship competences to establish the normative validity of assessment (Biesta, 2010). As elaborated in the fourth study of this dissertation types of assessment instruments may vary in the extent to which they facilitate different conceptions of citizenship. When choosing an assessment approach for citizenship competences it is therefore also relevant to take into account the specific conceptualization of citizenship competences.

Choosing a suitable assessment approach

Assessment of citizenship competences has proven no simple endeavor. The studies conducted as part of this dissertation have several implications for furthering the assessment of citizenship competences. Based on the studies conducted as part of this dissertation we reach several conclusions and

recommendations. These can be distinguished for the assessment of citizenship knowledge, attitudes, and skills respectively.

Assessment of citizenship knowledge

Students' citizenship knowledge can be assessed in different ways, and using rubrics proved to be moderately reliable. However, using rubrics for self-assessment showed that only one in four students provided an adequate explanation with the level they selected, severely limiting possibilities to make claims about their actual level. Consequently, it appears tests remain the most suitable instrument to assess citizenship knowledge. Notably, this refers not solely to multiple choice tests. While these have their benefits and applications, test items can be formulated on a spectrum from multiple choice items on the one end, to essay questions on the other. More open ended questions could also be graded using rubrics, provided students sufficiently elaborate their answers.

Assessment of citizenship attitudes

Most studies measuring citizenship competences use a set of Likert type items to measure citizenship attitudes. These items are easy to administer and generate a standardized outcome. The second and particularly the third study in this dissertation show citizenship attitudes can also be assessed using rubrics, and that such an approach can facilitate a comprehensive assessment of citizenship attitudes that more closely reflects the increasing complexity of more developed citizenship attitudes. There is no clear cut 'best instrument' to assess citizenship attitudes. Questionnaires offer many scales that have been tried and tested to provide standardized measurement of aspects of citizenship attitudes. Within the constraints of the construct-as-defined, questionnaires provide a very effective way to assess citizenship attitudes. If the purpose of assessment is to gain a richer picture of students' attitudes, rubrics clearly show promise.

Assessment of citizenship skills

Based on the explanations students provided to their self-assessed citizenship skills based on our rubrics, only one in four students could be considered to have provided an explanation that adequately reflected the selected level. This considerably limits any claims on their actual levels. Surveys using questionnaires to assess students' citizenship competences commonly assess 'citizenship self-efficacy', meaning students answer on a set of Likert type items to what extent they consider themselves able to perform certain tasks. Both self-assessment using rubrics or Likert type items could be prone to various biases such as the Dunning-

Kruger effect or students' assertiveness, clouding the measurement. We would welcome further studies into citizenship skills and their measurement, particularly approaches where students are enabled to perform authentic tasks based on which their skills can be assessed (see Chapter 5; cf. Roelofs & Sluijter, 2016).

Conclusion

This dissertation offers a synthesis of different approaches to assessing citizenship competences. Citizenship competences encompass a broad range of aspects, e.g. knowledge of democratic decision-making or attitudes towards societal differences. Assessment instruments therefore generally focus on a broad range of contents. Furthermore, citizenship competences can be assessed in different ways, and for different purposes. This dissertation *inter alia* presents a set of criteria based on which tests and questionnaires, portfolio, game-based assessment, and vignettes are reviewed. A newly developed set of rubrics to assess citizenship competences is tested and evaluated. Compared to the prevalent instruments that typically employ tests and questionnaires, rubrics offer a viable assessment that could help teachers and students gain a more comprehensive insight into their citizenship competences. The most fruitful way forward appears to invest in further expanding the instruments available, so that depending on the purpose of assessment the most suitable method of assessment may be selected.

APPENDIX A: RUBRICS FOR EXEMPLARY SOCIAL TASKS

Acting democratically

Knowledge (I know…)

Level A	Level B
You have general knowledge about what democracy and democratic are. You have an idea of how Dutch democracy works. You know democratic decision can also be made at home or at school. This level focuses on 'knowing'. *For example:* ■ *I know it's democratic to vote if not everyone wants the same thing.* ■ *I know the Netherlands has different political parties.* ■ *I know laws apply to everyone.*	You understand democracy happens in many places (nationally, but also in the classroom). You understand democracy implies certain rights and obligations. This level focuses on 'knowing' as well as 'understanding'. *For example:* ■ *I understand why it can be important to convince others when taking democratic decisions.* ■ *I know different political parties and can describe differences between them.* ■ *I understand why there are limits to the freedom of speech.*
Level C	Level D
You know how to apply your knowledge about democracy. You can explain how democratic processes in the Netherlands work. This level focuses on 'applying'. *For example:* ■ *I understand what consequences political decision can have, and what advantages and disadvantages that might have.* ■ *I can approach a (political) problem or discussion from different viewpoints.* ■ *I know there are different ways to reach a democratic decision.*	You can evaluate the Dutch democracy based on knowledge of democratic principles. You recognize the tensions involving democracy in practice. This level focuses on 'applying' as well as 'evaluating'. *For example:* ■ *I understand democratic decisions are made in different ways, and what advantages and disadvantages that might have.* ■ *I can explain when democratic decision-making is set up justly or unjustly.* ■ *I can evaluate what the advantages and disadvantages of a political decision (or proposal) might be in practice.*

Attitude (I believe…)

Level A	Level B
You believe people should be able to live together without problems. You think everyone should follow the rules. You think politics are irrelevant. *For example:* ■ *I believe everyone should be allow to provide input to a discussion.* ■ *I believe the minority should adjust to what the majority decides.* ■ *I believe there is too munch news about politics.*	You think it is important to act democratically. You think people should be considerate of others. You think politics are important, but you do not care much about it. *For example:* ■ *I believe some arguments weight heavier than others.* ■ *I believe I should listen to others, even if their opinion differs from mine.* ■ *I believe it is important to vote (when I turn 18).*
Level C	Level D
You want to have an open attitude and actively contribute to decisions. When taking decisions your take into account the feelings and interest of the minority. You show interest in national political, and try to keep up to date with the news. *For example:* ■ *I believe that counterarguments should be considered when taking a decision.* ■ *I am prepared to defend other people's interest, even when I don't agree with their position.* ■ *I believe everyone should regularly follow the news on political issues.*	You encourage others to have an open attitude. You are considerate of tensions and sensitive issues, and adjust your attitude accordingly. You feel involved with politics. *For example:* ■ *I believe it is important to consider consequences of (democratic) decisions for others, even when they are not involved in the decision making.* ■ *I believe it is important to make others aware of their part in democratic decisions.* ■ *I believe I should regularly follow the news, and develop my own opinion.*

Skills (I can…)

Level A	Level B
You can develop your own opinion and get your point across. *For example:* ■ *I can stick to my own opinion.* ■ *I can decide whether I'm in favor or against something.* ■ *I can stick to agreements we made together.*	You can take into account what others think in your daily environment. You can deal with disagreements and use arguments to convince others. *For example:* ■ *I can weight different opinions to consider my own viewpoint in a discussion.* ■ *I can formulate my opinion in a way that is not hurtful or insulting to others.* ■ *I can respectfully agree to disagree with someone.*
Level C	Level D
You can actively participate in discussions and do this in a thoughtful manner. You can adjust your own actions in discussions depending on the situation. *For example:* ■ *I can formulate a grounded opinion about current (political) events.* ■ *I can see matters from someone else's viewpoint when discussing controversial issues.* ■ *I can influence decisions made in my environment (at home, at school et cetera).*	You can adjust your own actions in discussions to the situation, even when it comes to complex or stressful situations. You are considerate of wider societal developments. *For example:* ■ *I can influence decisions made outside my environment (city council, political decisions et cetera).* ■ *I can bring a discussion to a conclusion that both proponents and opponents can get behind.* ■ *I can address sensitive or controversial topics in a discussion in a suitable way.*

Acting socially responsible

Knowledge (I know…)

Level A	Level B
You are familiar with social rules. You know what behavior benefits your environment, or damages you and others. This level focuses on 'knowing'. *For example:* ■ *I know I have to apologize if I offend someone.* ■ *I know why people have to pay taxes.* ■ *I know how I can be environmentally aware.*	You understand why it is important to interact with others and your environment in a favorable way. This level focuses on 'knowing' as well as 'understanding'. *For example:* ■ *I understand why there is a search for alternative energy (such as solar energy).* ■ *I understand why not everyone receives social security.* ■ *I understand why social cohesion at school and in society are important.*
Level C	**Level D**
You can explain how collective agreements (like taxes or social benefit) work. This level focuses on 'applying'. *For example:* ■ *I understand why knowledge of climate change is important.* ■ *I know what the 'welfare state' entails* ■ *I can explain what positive and negative consequences differences in wealth can have.*	You understand what tensions can arise in a diverse society, and how best to deal with those. You have knowledge of humanity's footprint on the environment. This level focuses on 'applying' as well as 'evaluating'. *For example:* ■ *I understand why countermeasures to climate change are difficult to establish.* ■ *I can evaluate whether social problems, such as homelessness or drug-use are dealt with in a favorable way.* ■ *I understand the tension between individual and collective interest.*

Attitude (I believe…)

Level A	Level B
You want to take others into account. You think people should take others into account in their behavior. *For example:* ◼ *I believe I should try not to offend other.* ◼ *I believe I should apologize when I know I'm in the wrong.* ◼ *I believe people should be able to safely live together.*	You want to help others in your area when they need help, and believe it is important people interact with each other favorably. You think everyone should contribute to 'living together'. *For example:* ◼ *I am prepared to help others, even when that doesn't help me.* ◼ *I believe it is important to show an interest in the people around m in my neighborhood* ◼ *I believe everyone should consider whether they want to be an organ donor.*
Level C	Level D
You want to actively contribute to your environment, and you make an effort to help others and your environment. You want to find a balance between your interests and those of others. *For example:* ◼ *I believe everyone should be aware of the impact of our actions on the climate.* ◼ *I try to promote the interests of others through volunteer work, class representative et cetera.* ◼ *I feel involved with society.*	You want to contribute to the way in which people behave towards others or the environment. You try to set an example. *For example:* ◼ *I want to make others aware of the world around them.* ◼ *I try, through my attitude, to encourage others to get involved.* ◼ *I try to offer counterbalance to opinions that can be considered hurtful or discriminating.*

Skills (I can…)

Level A	Level B
You can make sure you don't harm yourself or others in daily situations. *For example:* ■ *I can stand up for myself to friends or family.* ■ *I can take care of my own health.* ■ *I can live up to the trust others have in me.*	You can deal with others and the environment in a positive way in various situations. You can contribute to a positive social environment. *For example:* ■ *I can stand up for myself to others (teachers, strangers et cetera).* ■ *I can adapt my behavior to the environment.* ■ *I can give others the feeling I trust them.*
Level C	Level D
You can deal with others and the environment in a positive way in unfamiliar or new situations. You can assess the consequences of your behavior, and adjust your behavior accordingly. *For example:* ■ *I can stand up for the interests of others inside and outside my group.* ■ *I can address strangers respectfully to clean up their rubbish.* ■ *I can relate to others who are less fortunate than me, and indicate whether that is their own responsibility or if they deserve help.*	You relate your behavior to the wider community. You can purposefully adapt your behavior. You can make others aware of the consequences of their behavior. *For example:* ■ *I can give a well-founded opinion about climate change, and its implications for society.* ■ *I can make others aware of the importance to behave in a socially sustainable way.* ■ *I can make others aware of the importance to positively deal with their environment.*

Dealing with differences

Knowledge (I know…)

Level A	Level B
You know people can differ from each other (for example based on religion, culture, gender, social environment, lifestyle). You know the major differences between religions and cultures. This level focuses on 'knowing'. *For example:* ■ *I have general knowledge of the major religions (in the Netherlands).* ■ *I know what minority groups live in the Netherlands.* ■ *I can give examples of characteristics on which people differ.*	You have background knowledge about differences between people, and why those differences are relevant. You understand when something is sexist, discriminating, or racist. This level focuses on 'knowing' as well as 'understanding'. *For example:* ■ *I recognize intentional and unintentional discrimination.* ■ *I know cultural and societal differences between men and women.* ■ *I understand why it is important to take differences between people into account.*
Level C	Level D
You understand why people's background can be important to them. You understand what these differences mean to people and how to deal with those in a conscious way. This level focuses on 'applying'. *For example:* ■ *I understand the advantages and disadvantages that differences between people can have on group collaboration.* ■ *I understand the tension between equality and differences.* ■ *I understand how stressing differences can have positive and negative consequences.*	You know theories and background to (cultural) differences and can evaluate whether they are dealt with in a favorable way. This level focuses on 'applying' as well as 'evaluating'. *For example:* ■ *I recognize that people's opportunities are influenced by their social background.* ■ *I can explain how differences in society can have positive and negative consequences.* ■ *I can evaluate whether differences are dealt with in a favorable way in my environment.*

Attitude (I believe…)

Level A	Level B
You recognize differences between people, but who's different has to adapt. *For example:* ■ *I believe everyone should behave normally.* ■ *I believe people should adapt to the majority.* ■ *I believe I should treat everyone the same.*	You assume everyone should abide by the same rules. You believe that while people differ, everyone is equal. *For example:* ■ *I believe men and women deserve equal opportunity.* ■ *I believe people should be able to be who they want to be, as long as they abide by the rules.* ■ *I believe it's normal to respect other religion, sexes, or sexual preferences.*
Level C	Level D
You are open to differences and want to respect these. You believe it's important to give others the feeling they can be themselves around you. *For example:* ■ *I believe it's important that my behavior isn't insulting to others.* ■ *I'm open to learn about other cultures or customs.* ■ *I believe it's useful to consider my own prejudice.*	You show interest in differences and find it important to have a careful attitude towards how you deal with these. You are prepared to adjust and explain your attitude. *For example:* ■ *Through my attitude I try to express everyone deserves respect despite differences.* ■ *I believe it's important to consider varying viewpoints in a discussion.* ■ *I want to make others in my community aware of how they deal with differences.*

Skills (I can…)

Level A	Level B
You can treat others equally. You can behave in a way that doesn't hurt others. *For example:* ■ *I can behave in a way that isn't considered hurtful or insulting.* ■ *I can accept differences.* ■ *I can indicate my own boundaries.*	You can show compassion for other with a different background. You can deal with differences in daily situations by being considerate of others. *For example:* ■ *I can adjust my behavior to other people's customs or expectations.* ■ *I can estimate and respect other people's boundaries.* ■ *I can show respect for others.*
Level C	Level D
In situations where people from different backgrounds meet, you can behave in a way that is perceived positively. *For example:* ■ *I can look for ways to bridge differences.* ■ *I can utilize similarities and common interests.* ■ *I can deal with unfamiliar social situations.*	You can deal with situations where differences between people lead to tensions. You can view situations or events from different perspectives and have conversations about sensitive subjects. *For example:* ■ *I can act in a way that bridges differences between groups of people.* ■ *I can behave in a way that promotes an enjoyable contact in tense situations.* ■ *I can utilize the advantages of different perspectives.*

APPENDIX B: RUBRICS VOOR EXEMPLARISCHE SOCIALE TAKEN

Democratisch handelen

Kennis (weet…)

Niveau A	Niveau B
Je hebt algemene kennis over wat democratie en democratisch is. Je hebt een beeld van hoe de Nederlandse democratie werkt. Je weet dat democratische beslissingen ook thuis of op school genomen kunnen worden. Het gaat hier vooral om 'weten dat…'. *Bijvoorbeeld:* ■ *Ik weet dat het democratisch is om te stemmen als niet iedereen hetzelfde wil.* ■ *Ik weet dat er in Nederland verschillende politieke partijen zijn.* ■ *Ik weet dat wetten voor iedereen gelden.*	Je begrijpt dat democratie op allerlei manieren plaatsvindt (bijvoorbeeld landelijk, maar ook in de klas). Je begrijpt dat democratie bepaalde rechten en plichten schept. Het gaat hier om zowel 'weten' als 'begrijpen'. *Bijvoorbeeld:* ■ *Ik begrijp waarom het bij democratische beslissingen belangrijk kan zijn anderen te overtuigen.* ■ *Ik ken verschillende politieke partijen en kan verschillen daartussen aangeven.* ■ *Ik begrijp waarom er grenzen zijn aan de vrijheid van meningsuiting.*
Niveau C	Niveau D
Je weet hoe je jouw kennis over democratie in de praktijk kunt toepassen. Je kunt uitleggen hoe democratische processen in Nederland werken. Het gaat hier naast 'weten' en 'begrijpen' dus ook om 'toepassen'. *Bijvoorbeeld:* ■ *Ik begrijp welke gevolgen een politieke beslissing kan hebben, en welke voor- en nadelen aan een beslissing kleven.* ■ *Ik kan vanuit verschillende standpunten naar een (politiek) probleem of discussie kijken.* ■ *Ik weet dat er meerdere manieren zijn om tot een democratische beslissing te komen.*	Je kunt de Nederlandse democratie beoordelen vanuit je kennis van democratische principes. Je herkent spanningen rond democratie in de praktijk. Het gaat hier zowel om 'toepassen' als 'evalueren'. *Bijvoorbeeld:* ■ *Ik begrijp dat democratische beslissingen op verschillende manieren genomen kunnen worden, en welke voor- en nadelen dat heeft.* ■ *Ik kan uitleggen wanneer democratische besluitvorming wel of niet eerlijk is ingericht.* ■ *Ik kan evalueren wat de voor- en nadelen zijn van een politieke beslissing (of voorstel) in de praktijk.*

Attitude (vind…)

Niveau A	Niveau B
Je vindt dat mensen zonder problemen met elkaar om moeten kunnen gaan. Je vindt dat iedereen zich aan de regels moet houden. Je vindt politiek niet relevant. *Bijvoorbeeld:* ■ *Ik vind dat iedereen input mag leveren aan een discussie.* ■ *Ik vind dat de minderheid zich moet aanpassen aan wat de meerderheid beslist.* ■ *Ik vind dat er teveel nieuws is over de politiek.*	Je vindt het belangrijk om je democratisch te gedragen. Je vindt dat mensen rekening moeten houden met elkaar. Je vindt de politiek wel belangrijk, maar interesseert je er weinig voor. *Bijvoorbeeld:* ■ *Ik vind dat sommige argumenten zwaarder wegen dan andere.* ■ *Ik vind dat ik naar anderen hoor te luisteren, ook als zij een andere mening hebben dan ik.* ■ *Ik vind het belangrijk om (vanaf mijn 18^e) te gaan stemmen.*
Niveau C	Niveau D
Je wilt een open houding hebben en actief meedenken over beslissingen. Je houdt bij beslissingen rekening met de positie en gevoelens van de minderheid. Je toont interesse in landelijke politiek, en probeert op de hoogte te blijven van het nieuws daarover. *Bijvoorbeeld:* ■ *Ik vind dat bij het nemen van beslissingen nagedacht moet worden over tegenargumenten.* ■ *Ik ben bereid de belangen van anderen te verdedigen, ook als ik het inhoudelijk niet met ze eens ben.* ■ *Ik vind dat iedereen regelmatig het nieuws over de politiek moet volgen.*	Je stimuleert bij anderen een open houding. Je hebt oog voor spanningen en gevoelige situaties, en kunt jezelf daarin goed een houding geven. Je voelt je betrokken bij de politiek. *Bijvoorbeeld:* ■ *Ik vind het belangrijk om bij (democratische) beslissingen na te denken over de gevolgen voor anderen, ook als zij geen inspraak hebben.* ■ *Ik vind het belangrijk anderen na te laten denken over hun rol in democratische beslissingen.* ■ *Ik vind dat ik regelmatig het politieke nieuws moet volgen en daar ook een eigen mening en visie over vormen.*

Vaardigheid (kan…)

Niveau A	Niveau B
Je kunt je eigen mening vormen en jouw standpunt duidelijk maken. *Bijvoorbeeld:* ■ *Ik kan vasthouden aan mijn mening.* ■ *Ik kan bepalen of ik ergens voor of tegen ben.* ■ *Ik kan me houden aan afspraken die we gezamenlijk gemaakt hebben.*	Je kunt rekening houden met wat anderen in je dagelijkse omgeving vinden. Je kunt omgaan met meningsverschillen en anderen met argumenten overtuigen. *Bijvoorbeeld:* ■ *Ik kan op basis van verschillende meningen een eigen standpunt innemen in discussies.* ■ *Ik kan mijn mening zo formuleren dat ik anderen niet beledig of kwets.* ■ *Ik kan een discussie waarin we het niet eens worden goed afsluiten.*
Niveau C	Niveau D
Je kunt actief deelnemen aan discussies en doet dat op een doordachte manier. Je kunt je eigen handelen in discussies aanpassen aan de situatie. *Bijvoorbeeld:* ■ *Ik kan een onderbouwde mening vormen over actuele (politieke) gebeurtenissen.* ■ *Ik kan mij bij het bespreken van controversiële onderwerpen inleven in het standpunt van een ander.* ■ *Ik kan invloed hebben op beslissingen in mijn omgeving (zoals op school, thuis etc.).*	Je kunt je eigen handelen in discussies aanpassen aan de situatie, ook wanneer het om complexe of spanningsvolle situaties gaat. Je hebt oog voor bredere maatschappelijke ontwikkelingen. *Bijvoorbeeld:* ■ *Ik kan invloed hebben op beslissingen buiten mijn omgeving (zoals in de gemeente of politiek).* ■ *Ik kan in discussies een conclusie bereiken waar voor- en tegenstanders achter kunnen staan.* ■ *Ik kan in een discussie gevoelige of controversiële onderwerpen op een goede manier aankaarten.*

Maatschappelijk verantwoord handelen

Kennis (weet…)

Niveau A	Niveau B
Je kent algemene omgangsnormen. Je weet welk gedrag bijdraagt aan je omgeving, of anderen of jezelf juist schaadt of onrecht aandoet. Het gaat hier vooral om 'weten'. *Bijvoorbeeld:* ■ *Ik weet dat ik mijn excuses moet aanbieden als ik iemand beledigd heb.* ■ *Ik weet waarom mensen belasting betalen.* ■ *Ik weet hoe ik milieubewust kan zijn.*	Je begrijpt waarom het belangrijk is op een goede manier met anderen en je omgeving om te gaan. Het gaat hier zowel om 'weten' als 'begrijpen'. *Bijvoorbeeld:* ■ *Ik begrijp waarom gezocht moet worden naar alternatieve energiebronnen (zoals zonne-energie).* ■ *Ik begrijp waarom niet iedereen een uitkering krijgt.* ■ *Ik begrijp waarom samenhang op school en in de maatschappij belangrijk is.*
Niveau C	Niveau D
Je kunt uitleggen hoe collectieve afspraken (zoals belasting en uitkeringen) werken. Het gaat hier naast 'weten' en 'begrijpen' dus ook om 'toepassen'. *Bijvoorbeeld:* ■ *Ik begrijp waarom bewustwording van klimaatverandering belangrijk is.* ■ *Ik weet wat 'de verzorgingsstaat' inhoudt.* ■ *Ik kan uitleggen welke positieve en negatieve effecten verschillen in welvaart (rijkdom) kunnen hebben.*	Je begrijpt welke spanningen kunnen ontstaan in een diverse samenleving, en hoe daar mee omgegaan kan worden. Je hebt kennis over de impact van de mens op het milieu. Het gaat hier zowel om 'toepassen' als 'evalueren'. *Bijvoorbeeld:* ■ *Ik begrijp waarom maatregelen tegen klimaatverandering vaak moeilijk tot stand komen.* ■ *Ik kan evalueren of maatschappelijke problemen, zoals dakloosheid of drugsgebruik, op een goede manier aangepakt worden.* ■ *Ik begrijp de spanning tussen individueel belang en algemeen belang.*

Attitude (vind…)

Niveau A	Niveau B
Je wilt rekening houden met anderen. Je vindt dat mensen in hun gedrag rekening moeten houden met anderen. *Bijvoorbeeld:* ■ *Ik vind dat ik moet proberen om anderen niet te beledigen.* ■ *Ik vind dat ik excuses moet kunnen aanbieden wanneer ik weet dat ik fout zit.* ■ *Ik vind dat mensen veilig naast elkaar moeten kunnen leven.*	Je wilt anderen in je omgeving helpen wanneer zij hulp nodig hebben, en vindt het belangrijk dat mensen op een goede manier met elkaar omgaan. Je vindt dat iedereen moet bijdragen aan 'samen-leven'. *Bijvoorbeeld:* ■ *Ik ben bereid anderen te helpen, ook als ik daar zelf niet iets aan heb.* ■ *Ik vind het van belang om belangstelling te tonen voor mensen in mijn buurt/wijk.* ■ *Ik vind dat iedereen moet nadenken of hij/zij wel of geen orgaandonor wil zijn.*
Niveau C	Niveau D
Je wilt een actieve bijdrage leveren aan de leefomgeving, en zet je in voor anderen en je omgeving. Je wilt een evenwicht vinden tussen je eigen belangen en die van anderen. *Bijvoorbeeld:* ■ *Ik vind dat iedereen zich bewust moet zijn van de effecten van ons gedrag op het klimaat.* ■ *Ik probeer mij als vrijwilliger, klassenvertegenwoordiger etc. in te zetten voor de belangen van anderen.* ■ *Ik voel mij betrokken bij de samenleving.*	Je wilt een bijdrage leveren aan de manier waarop anderen met elkaar en met de omgeving omgaan. Je probeert een voorbeeldrol te vervullen. *Bijvoorbeeld:* ■ *Ik wil anderen in mijn omgeving bewust maken van de wereld om hen heen.* ■ *Ik probeer met mijn houding de betrokkenheid van anderen te stimuleren.* ■ *Ik probeer tegenwicht te bieden aan meningen die kwetsend of discriminerend kunnen zijn.*

Vaardigheid (kan…)

Niveau A	Niveau B
Je kunt in alledaagse situaties ervoor zorgen dat je voor jezelf zorgt en jezelf geen schade doet . *Bijvoorbeeld:* ■ *Ik kan voor mezelf opkomen bij vrienden/bekenden.* ■ *Ik kan zorgen voor mijn eigen gezondheid.* ■ *Ik kan het vertrouwen dat anderen in mij hebben waarmaken.*	Je kunt in verschillende situaties op een positieve manier omgaan met anderen en de omgeving. Je kunt bijdragen aan een positieve sociale omgeving. *Bijvoorbeeld:* ■ *Ik kan opkomen voor mijn belangen bij anderen (docenten, op straat etc.).* ■ *Ik kan in mijn gedrag rekening houden met het milieu.* ■ *Ik kan anderen het gevoel geven dat ik ze vertrouw.*
Niveau C	Niveau D
Je kunt ook in onbekende of nieuwe situaties je gedrag op een doordachte manier afstemmen op anderen. Je kunt de gevolgen van je gedrag inschatten en je gedrag daarop aanpassen. *Bijvoorbeeld:* ■ *Ik kan opkomen voor de belangen van anderen binnen en buiten mijn groep.* ■ *Ik kan onbekenden respectvol aanspreken om hun rommel op te ruimen.* ■ *Ik kan mij inleven in de situatie van anderen die het minder goed hebben, en aangeven of dat hun eigen verantwoordelijkheid is of dat zij ondersteuning verdienen.*	Je relateert je eigen gedrag ook aan de bredere gemeenschap. Je kunt je gedrag afstemmen om een doel te bereiken. Je kunt anderen bewust maken van de gevolgen van hun gedrag. *Bijvoorbeeld:* ■ *Ik kan een onderbouwde mening vormen over klimaatverandering, en de implicaties daarvan voor de samenleving.* ■ *Ik kan anderen bewust maken van het belang om op een duurzame manier met elkaar om te gaan.* ■ *Ik kan anderen bewust maken van het belang om goed met de omgeving om te gaan.*

Omgaan met verschillen

Kennis (weet…)

Niveau A	Niveau B
Je weet dat mensen van elkaar kunnen verschillen (zoals geloof, cultuur, geslacht, sociaal milieu, leefwijze e.d.). Je weet wat de belangrijkste verschillen zijn tussen godsdiensten en tussen culturen. Het gaat hier vooral om 'weten'. *Bijvoorbeeld:* ■ *Ik heb algemene kennis van de grootste religies (in Nederland).* ■ *Ik weet welke minderheidsgroepen er in Nederland zijn.* ■ *Ik kan voorbeelden noemen van punten waarop mensen van elkaar verschillen.*	Je kent de achtergrond van verschillen tussen mensen en waarom die verschillen relevant zijn. Je begrijpt wanneer iets seksistisch, discriminerend of racistisch is. Het gaat hier om zowel 'weten' als 'begrijpen'. *Bijvoorbeeld:* ■ *Ik herken bedoelde en onbedoelde discriminatie.* ■ *Ik ken culturele en maatschappelijke verschillen tussen mannen en vrouwen.* ■ *Ik begrijp waarom het belangrijk is rekening te houden met verschillen tussen mensen.*
Niveau C	Niveau D
Je begrijpt waarom de achtergrond van mensen voor hen belangrijk kan zijn. Je begrijpt wat deze verschillen voor mensen betekenen en hoe je daar op een goede manier mee om kunt gaan. Het gaat hier dus om 'toepassen'. *Bijvoorbeeld:* ■ *Ik begrijp welke voor- of nadelen verschillen tussen mensen kunnen hebben voor het functioneren van een groep.* ■ *Ik begrijp de spanning tussen gelijkheid en verschillen.* ■ *Ik begrijp hoe het benadrukken van verschillen positieve en negatieve gevolgen kan hebben.*	Je kent theorieën en achtergronden bij (culturele) verschillen en kunt evalueren of daar op een goede manier mee omgegaan wordt . Het gaat hier zowel om 'toepassen' als 'evalueren'. *Bijvoorbeeld:* ■ *Ik herken dat de kansen die mensen hebben beïnvloed worden door hun sociale achtergrond.* ■ *Ik kan uitleggen hoe verschillen in de samenleving positieve of negatieve gevolgen kunnen hebben.* ■ *Ik kan evalueren of in mijn omgeving op een goede manier met verschillen omgegaan wordt.*

Attitude (vind…)

Niveau A	Niveau B
Je ziet verschillen tussen mensen, maar vindt dat wie anders is zich moet aanpassen. *Bijvoorbeeld:* ■ *Ik vind dat iedereen zich normaal moet gedragen.* ■ *Ik vind dat mensen zich moeten aanpassen aan de meerderheid.* ■ *Ik vind dat ik iedereen hetzelfde hoor te behandelen.*	Je gaat er van uit dat iedereen zich aan dezelfde regels en afspraken moet houden. Je vindt dat wanneer mensen verschillend zijn, iedereen evenveel waard is. *Bijvoorbeeld:* ■ *Ik vind dat mannen en vrouwen dezelfde kansen verdienen.* ■ *Ik vind dat er ruimte moet zijn om jezelf te zijn als je je aan de regels houdt.* ■ *Ik vind het normaal om andere godsdiensten, geslachten of seksuele geaardheden te respecteren.*
Niveau C	**Niveau D**
Je staat open voor verschillen en wilt daar ook rekening mee houden. Je vindt het belangrijk anderen het gevoel te geven dat ze zichzelf kunnen zijn waar jij bij bent. *Bijvoorbeeld:* ■ *Ik vind het belangrijk dat mijn gedrag niet beledigend is voor anderen.* ■ *Ik sta open om over andere culturen of gebruiken te leren.* ■ *Ik vind het zinvol om na te denken over mijn eigen vooroordelen.*	Je toont interesse in verschillen en vindt het belangrijk om een zorgvuldige houding te hebben over hoe je daarmee omgaat. Je bent bereid die houding uit te leggen en bij te stellen. *Bijvoorbeeld:* ■ *Ik probeer met mijn houding uit te dragen dat iedereen respect verdient ondanks verschillen.* ■ *Ik vind het belangrijk dat discussies vanuit verschillende perspectieven bekeken worden.* ■ *Ik wil anderen in mijn omgeving bewust maken van hoe zij met verschillen omgaan.*

Vaardigheid (kan…)

Niveau A	Niveau B
Je kunt anderen gelijk behandelen. Je kunt je op een manier gedragen die anderen niet schaadt. *Bijvoorbeeld:* ■ *Ik kan mij gedragen op een manier die niet als kwetsend of beledigend opgevat wordt.* ■ *Ik kan verschillen accepteren.* ■ *Ik kan mijn eigen grenzen aangeven.*	Je kunt je inleven in anderen met een andere achtergrond. Je kunt in dagelijkse situaties omgaan met verschillen door rekening te houden met anderen. *Bijvoorbeeld:* ■ *Ik kan mijn gedrag aanpassen aan andermans wensen of gewoonten.* ■ *Ik kan de grenzen van anderen inschatten en respecteren.* ■ *Ik kan respect tonen voor anderen.*
Niveau C	Niveau D
Je kunt in situaties waarin mensen met verschillende achtergronden samenkomen je op een manier gedragen die door iedereen als positief wordt ervaren. *Bijvoorbeeld:* ■ *Ik kan naar manieren zoeken om verschillen te overbruggen.* ■ *Ik kan overeenkomsten en gemeenschappelijke interesses benutten.* ■ *Ik kan omgaan met onbekende sociale situaties.*	Je kunt omgaan met situaties waarin verschillen tussen mensen tot spanningen leiden. Je kunt situaties en gebeurtenissen vanuit verschillende perspectieven bekijken en gesprekken voeren over onderwerpen die gevoelig zijn. *Bijvoorbeeld:* ■ *Ik kan activiteiten ondernemen waardoor verschillen tussen groepen overbrugd worden.* ■ *Ik kan in gespannen situaties bijdragen aan een plezierige omgang tussen mensen.* ■ *Ik kan de voordelen van verschillende perspectieven benutten.*

APPENDIX C: REFLECTION ASSIGNMENTS

Reflection assignment acting socially responsible

In the previous lesson you've played the social ladder game. Based on your role, you made it more of less far on the social ladder. In this assignment we delve further into the meaning of social differences in society.

- Complete the assignment on the computer.
- Individual assignment.
- Questions 2 to 8 are about your opinion, not from the perspective you took during the game.
- Due date: ..
- Requirements: at least 1 page, maximum 3 pages.

Questions:

1. What role did you have during the social ladder game?
2. Did you feel you got far enough up the ladder? Why?
3. What did you think about the differences in steps for the different roles?
4. a) Do you so those differences in Dutch society? Are those differences large?
   b) Do you think those differences are fair? Explain <u>why</u>.
5. What positive consequences can social differences have?
6. What negative consequences can social differences have?
7. How do you think the government should deal with people on the top or bottom of the social ladder? Explain why.

Reflection assignment dealing with differences

Substantive reflection:

- ➤ Describe the differences and similarities between your own culture and the culture you chose for the assignment.
- ➤ Are there prejudices about this group of immigrants in the Netherlands? How can you tell? Explain your answer using examples.
- ➤ Respond to the following statements:
    - o Your background / ethnicity has consequences for your chance of success in society.
    - o Instead of looking at differences between cultures, we should be looking at similarities.
    - o Some cultures are better than others.

REFERENCES

Abdelzadeh, A., & Lundberg, E. (2017). Solid or flexible? Social trust from early adolescence to young adulthood. *Scandinavian Political Studies, 40*(2), 207-227. http://dx.doi.org/10.1111/1467-9477.12080.

Ainley, J., Schulz, W., & Friedman, T. (2013). *ICCS 2009 encyclopedia: Approaches to civic and citizenship education around the world*. Amsterdam, Netherlands: IEA.

Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). Berkshire, England: Open University Press.

Aldenmyr, S. I., Wigg, U. J., & Olson, M. (2012). Worries and possibilities in active citizenship: Three Swedish educational contexts. *Education, Citizenship and Social Justice, 7*(3), 255-270.

Allport, G. W. (1935). Attitudes. In C. Murchison (ed.), *Handbook of social psychology* (798–844). Worcester, MA: Clark University Press.

Almgren, E. (2006). *Att fostra demokrater: Om skolan i demokratin och demokratin i skolan [Fostering democrats: About school in democracy and democracy in school]*. Uppsala: Acta Universitatis Upsaliensis.

Alwin, D., & Krosnick, J. (1991). Aging, cohorts, and the stability of sociopolitical orientations over the life span. *American Journal of Sociology, 97*(1), 169-195. http://www.jstor.org/stable/2781642.

American Psychological Association. (1997). *Learner-centered psychological principles: A framework for school reform and redesign*. Washington, DC: American Psychological Association.

Amnå, E. (2012). How is civic engagement developed over time? Emerging answers from a multidisciplinary field. *Journal of Adolescence, 35*, 611–627. http://dx.doi.org/10.1016/j.adolescence.2012.04.011.

Amnå, E., Ekström, M., Kerr, M., & Stattin, H. (2009). Political socialization and human agency: The development of civic engagement from adolescence to adulthood. *Statsvetenskaplig Tidskrift, 111*, 27–40.

Amnå, E., Ekman, T., & Almgren, E. (2007). The End of a Distinctive Model of Democracy? Country-diverse Orientations among Young Adult Scandinavians. *Scandinavian Political Studies, 30*(1), 61-86.

Amnå, E., Englund, T., & Ljunggren, C. (2010). Skolor som politiska arenor [The school as political arena]. In Skolverket, *Skolor som politiska arenor* (pp. 8-16). Stockholm: Fritzes kundservice.

Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53*(1), 27-30. http://dx.doi.org/10.3200/CTCH.53.1.27-31.

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159-181. http://dx.doi.org/10.1080/02602930600801928.

Andrews, R., & Mycock, A. (2007). Citizenship education in the UK: Divergence within a multinational state. *Citizenship Teaching and Learning, 3*(1), 73-88.

Arensmeier, C. (2010). The democratic common sense. Young Swedes' understanding of democracy - theoretical features and educational incentives. *Young - Nordic Journal of Youth Research, 18*(2), 197-222.

Arnesen, A.L., & Lundahl, L. (2006). Still social and democratic? Inclusive education policies in the Nordic welfare states. *Scandinavian Journal of Educational Research, 50*(3), 285-300.

Atzmüller, C., & Steiner, P. (2010). Experimental vignette studies in survey research. *Methodology, 6*, 128–138. http://dx.doi.org/10.1027/1614-2241/a000014.

Auspurg, K., Hinz, T., & Liebig, S. (2009). *Complexity, learning effects, and plausibility of vignettes in factorial surveys*. Paper presented at the American sociological association annual meeting. http://nbn-resolving.de/urn:nbn:de:bsz:352-150806.

Bakken, A., & Elstad, J. (2012). *For store forventninger? Kunnskapsløftet og ulikhetene i grunnskolekarakterer [Great expectations? The Knowledge Promotion reform and inequalities in primary grades]*. Oslo, Norway: NOVA – Norwegian Social Research.

Barber, B. (1984). *Strong democracy. Participatory politics for a new age*. Berkeley, CA: University of California Press.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction, 2013*, 1–11. http://dx.doi.org/10.1155/2013/136864.

Berman, S. (1990). Educating for social responsibility. *Educational Leadership, 43*(3), 75-80.

Bernitz, H. (2012). *Country Report: Sweden*. Badia Fiesolana, Italy: EUDO Citizenship Observatory.

Bernmark-Ottosson, A. (2005). *Demokratins Stöttepelare: En studie av lärarstuderandes demokratiuppfattningar [Anchors of democracy: A study*

of democracy conceptions]. Karlstad: Institutionen for utbildningvetenskap.

Biesta, G. (2009). What kind of citizen? What kind of democracy? Citizenship education and the Scottish Curriculum for Excellence. *European Educational Research Journal, 8*(2), 146–157.

Biesta, G. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. London: Paradigm Publishers.

Biseth, H. (2009). Democracy and education in a multicultural Scandinavia: What mandate is designated to educators? *Intercultural Education, 20*(3), 243-254.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7-74. http://dx.doi.org/10.1080/0969595980050102.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: McKay.

Borhaug, K. (2006). Mission impossible? School level student democracy. *Citizenship, Social and Economic Education, 7*(1), 26-41.

Borhaug, K. (2010). Norwegian civic education - Beyond formalism? *Journal of Social Science Education,* 9(1), 66-77.

Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge, England: University of Cambridge, School of Education.

Bron, J. (2006). *Een basis voor burgerschap: een inhoudelijke verkenning voor het funderend onderwijs [A basis for citizenship: A content exploration for foundational education]*. Enschede: Stichting Leerplanontwikkeling (SLO).

Bron, J., & Thijs, A. (2010). Het leerplan burgerschap in ontwikkeling [The curriculum for citizenship in development]. In J. L. Peschar, H. Hooghoff, A. B. Dijkstra, & G. ten Dam (Eds.), *Scholen voor burgerschap. Naar een kennisbasis voor burgerschapsonderwijs [Scholing for citizenship. To a knowledge base for citizenship education]*. Apeldoorn: Garant.

Bron, J., & Thijs, A. (2011). Leaving it to the schools: Citizenship, diversity and human rights education in schools. *Educational Research, 53*(2), 123-136.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronneman-Helmers, R., & Zeijl, E. (2008). Burgerschapsvorming in het onderwijs [Citizenship development in education]. In P. Schnabel, R. Bijl, & J. d. Hart, *Betrekkelijke betrokkenheid. Studies in sociale cohesie*

[Relative involvement. Studies into social cohesion] (pp. 173-208). Den Haag: SCP.

Carr, D. (2003). Values education in Scotland. In T. Bryce, & W. Humes (Eds.), *Scottish Education: Post-devolution* (2nd ed., pp. 323-330). Edinburgh: University Press.

CBS. (2016). *Annual report on integration*. The Hague, Netherlands: Statistics Netherlands.

Chin, J., Dukes, R., & Gamson, W. (2009). Assessment in simulation and gaming: A review of the last 40 years. *Simulation & Gaming, 40*, 553–568. http://dx.doi.org/10.1177/1046878109332955.

Christophersen, J., Lotsberg, D. O., Knutsen, K., & Borhaug, K. (2003). *Evaluering av samfunnsfag i Reform97 Syntheserapport [Evaluation of social studies in the 97 reform]*. Bergen: Hogskolen i Bergen.

Clogg, C. C., & Eliason, S. R. (1987). Some common problems in log-linear analysis. *Sociological Methods & Research, 16*(1), 8-44. http://dx.doi.org/10.1177/0049124187016001002.

Corti, K. (2006). *Games-based learning; a serious business application*. PIXE Learning Limited.

Council of Europe. (2005). *All-European study on education for democratic citizenship policies*. Strasbourg: Council of Europe.

Council of Europe. (2010). *Charter on education for democratic citizenship and human rights education*. Strasbourg, France: Council of Europe.

Council of Europe. (2016). *Competences for democratic culture: Living together as equals in culturally diverse democratic societies*. Strasbourg, France: Council of Europe.

Cowan, E., & McMurtry, D. (2009). The implementation of 'Education for Citizenship' in schools in Scotland: A research report. *The Curriculum Journal, 20*(1), 61-72.

Crick, B. (2007). Citizenship: The political and the democratic. *British Journal of Educational Studies, 55*(3), 235-248. http://www.jstor.org/stable/4620567.

Cross, B., Hall, J., Hall, S., Hulme, M., Lewin, J., & McKinney, S. (2009). *Pupil Participation in Scottish Schools: Final Report. Learning and Teaching Scotland*.

Dahl, R. A. (1998). *On democracy*. New Haven, CT: Yale University Press.

Davies, A., & LeMahieu, P. (2003). Assessment for learning: Reconsidering portfolios and research evidence. In M. Segers, & F. Dochy (Eds.), *Optimising new modes of assessment: In search of quality standards* (pp.

141–170).Dordrecht, Netherlands: Kluwer Academic Publishers. http://dx.doi.org/10.1007/0-306-48125-1_7.

De Groot, I. (2011). Why we are not democratic yet: The complexity of developing a democratic attitude. In W. Veugelers (ed.), *Education and humanism: Linking autonomy and humanity* (79-94). Rotterdam, Netherlands: Sense Publishers.

Deuchar, R. (2003). Preparing tomorrow's people: The new challenges of citizenship education for involving Scottish pupils and teachers in participative decision-making processes. *Scottish Educational Review, 35*(1), 27-37.

Deuchar, R. (2009). Seen and heard, and then not heard: Scottish pupils' experience of democratic educational practice during the transition from primary to secondary school. *Oxford Review of Education, 35*(1), 23-40.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation, 27*(4), 307-329. http://dx.doi.org/10.1016/S0191-491X(01)00032-3.

Dieterle, E., & Clarke, J. (2008). Multi-user virtual environments for teaching and learning. In M. Pagani (Ed.), *Encyclopedia of multimedia technology and networking* (pp. 1033–1041). Hershey, PA: Idea Group, Inc.

Dijkstra, A. B., & de la Motte, P. I. (Eds.). (2014). *Social outcomes of education: The assessment of social outcomes and school improvement through school inspections*. Amsterdam, Netherlands: Amsterdam University Press.

Dijkstra, A. B., de la Motte, P. I., & Eilard, A. (2014a). Social outcomes of education. Concept and measurement. In A. B. Dijkstra, & P. I. de la Motte (Eds.), *Social outcomes of education: The assessment of social outcomes and school improvement through school inspections* (pp. 29–49).Amsterdam, Netherlands: Amsterdam University Press.

Dijkstra, A. B., de la Motte, P. I., Gursli, R., Maxwell, S., Dahlberg, B., El Khayati, N., & Vosse, A. (2014b). Evaluating social outcomes of schools. Models of assessment. In A. B. Dijkstra, & P. I. delaMotte (Eds.), *Social outcomes of education: The assessment of social outcomes and school improvement through school inspections* (pp. 171–188). Amsterdam, Netherlands: Amsterdam University Press.

Directorate of Integration and Diversity. (2009). *Overview 2008: How well is integration working?* Oslo: Directorate of Integration and Diversity.

Dochy, F., & Moerkerke, G. (1997). Assessment as a major influence on learning and instruction. *International Journal of Educational Research, 27*, 415–432.

Dochy, F. (2001). Education assessment: Major developments. In N. Smelser, & P. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 4249–4255). Amsterdam, Netherlands: Elsevier. http://dx.doi.org/10.1016/B0-08-0430767/02367-6.

Doppen, F. H. (2010). Citizenship education and the Dutch national identity debate. *Education, Citizenship and Social Justice, 5*(2), 131-143.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237-251. http://dx.doi.org/10.3102/0013189X15584327.

Eckstein, H. (1966). *A theory of stable democracy*. Princeton, NJ: Princeton University Press.

Eckstein, K., Jugert, P., Noack, P., Born, M., & Sener, T. (2015). Comparing correlates of civic engagement between immigrant and majority youth in Belgium, Germany, and Turkey. *Research in Human Development*, *12*(1-2), 44-62. http://dx.doi.org/10.1080/15427609.2015.1010346.

Eckstein, K., Noack, P., & Gniewosz, B. (2012). Attitudes toward political engagement and willingness to participate in politics: Trajectories throughout adolescence. *Journal of Adolescence, 35*, 485–495. http://dx.doi.org/10.1016/j.adolescence.2011.07.002.

Education Scotland. (2013a). *Curriculum for Excellence: Political literacy. CfE Briefing 14*. Glasgow: Education Scotland.

Education Scotland. (2013b). *Political literacy and the referendum on independence for Scotland*. Glasgow: Education Scotland.

Education Scotland. (2013c). *Social Studies 3-18*. Livingston: Education Scotland.

Eichhorn, J. (2013). *Will 16 and 17 year olds make a difference in the referendum?* Edinburgh: ScotCen Social Research.

Eidhof, B., ten Dam, G., Dijkstra, A. B., & van de Werfhorst, H. (2017). Youth citizenship at the end of primary school: the role of language ability. *Research Papers in Education, 32*(2), 217-230. http://dx.doi.org/10.1080/02671522.2016.1167235.

Ekman, J., & Amnå, E. (2012). Political participation and civic engagement: Towards a new typology. *Human Affairs, 22*, 283–300. http://dx.doi.org/10.2478/s13374-012-0024-1.

Eriksson, C. (2006). *Det borde vara att folket bestämmer [It should be that the people decide]*. Orebro: Orebro University.

Eseryel, D., Ifenthaler, D., & Ge, X. (2011). Alternative assessment strategies for complex problem solving in game-based learning environments. In D. Ifenthaler, P. Kinshuk, D. Isaias, M. Sampson, & G. Spector (Eds.), *Multiple perspectives on problem solving and learning in the digital age* (pp. 159–178). Brussels, Belgium: Eurydice. http://dx.doi.org/10.1007/978-1-4419-7612-3_11.

Eurydice. (2012). *Citizenship education in Europe*. Brussels: Eurydice.

Eurydice. (2017). *Citizenship education at school in Europe – 2017*. Luxembourg: Publications Office of the European Union.

Fischman, G. E., & Haas, E. (2012). Beyond idealized citizenship education: Embodied cognition, metaphors, and democracy. Review of Research in Education, 36, 169– 196. http://dx.doi.org/10.3102/0091732X11420927.

Fjeldstad, D., & Mikkelsen, R. (2003). Strong democratic competence does not automatically lead to strong engagement and participation. I*nternational Journal of Educational Research, 39*, 621-632.

Fraile, J., Panadero, E. & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation, 53*, 69-76. http://dx.doi.org/10.1016/j.stueduc.2017.03.003.

Freeman, M. (2009). Editorial: Education and citizenship in modern Scotland. *History of Education, 38*(3), 327-332.

Friedman, M. (1955). The Role of Government in Education. In R. A. Solo (ed.), *Economics and the public interest*. New Brunswick, NJ: Rutgers University Press.

Galston, W. (2001). Political knowledge, political engagement, and civic education. *Annual Review of Political Science, 4*(1), 217-234

Geboers, E., Geijsel, F., Admiraal, W., Jorgensen, T., & Ten Dam, G. (2015). Citizenship development of adolescents during the lower grades of secondary education. *Journal of Adolescence, 45*, 89-97. http://dx.doi.org/10.1016/j.adolescence.2015.08.017.

Geboers, E., Geijsel, F., Admiraal, W., & Ten Dam, G. (2013). Review of the effects of citizenship education. *Educational Research Review, 9*, 158-173. http://dx.doi.org/10.1016/j.edurev.2012.02.001.

Geboers, E., Geijsel, F., Admiraal, W., & Ten Dam, G. (2015). Citizenship orientations and knowledge in primary and secondary education. *Social Psychology of Education*. http://dx.doi.org/10.1007/s11218-014-9265-7.

Geijsel, F., Ledoux, G., Reumerman, R., & Ten Dam, G. (2012). Citizenship in young people's daily lives: differences in citizenship competences of adolescents in the Netherlands. *Journal of Youth Studies, 15*, 711-729. http://dx.doi.org/10.1080/13676261.2012.671932.

Gordon, E., & Baldwin-Philippi, J. (2014). Playful civic learning: Enabling reflection and lateral trust in game-based public participation. *International Journal of Communication, 8*, 759–786.

Gulikers, J., Biemans, H., & Mulder, M. (2009). Developer, teacher, student and employer evaluations of competence-based assessment quality. *Studies in Educational Evaluation, 35*, 110–119. http://dx.doi.org/10.1016/j.stueduc.2009.05.002.

Hahn, C. (1999). Citizenship education: An empirical study of policy, practices and outcomes, *Oxford Review of Education, 25*(1-2), 231-250. http://dx.doi.org/10.1080/030549899104233.

Hainmueller, J., Hangertner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences, 112*, 2395–2400. http://dx.doi.org/10.1073/pnas.1416587112.

Hanghøj, T. (2011). Clashing and emerging genres: The interplay of knowledge forms in educational gaming. *Design for Learning, 4*, 22–33.

Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal, 16*, 207–223. http://dx.doi.org/10.1080/09585170500136093.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. http://dx.doi.org/10.3102/003465430298487.

Hilbers, G., Dekkers, H., & Dijkstra, A. B. (2010). De ontwikkeling van burgerschapsonderwijs op scholen voor primair en voortgezet onderwijs [The development of citizenship education in schools for primary and secondary education]. In J. L. Peschar, H. Hooghoff, A. B. Dijkstra, & G. ten Dam (Eds.), *Scholen voor burgerschap. Naar een kennisbasis voor burgerschapsonderwijs [Schooling for citizenship. To a knowledge base for citizenship education]*. Apeldoorn: Garant.

HMIE. (2006a). *Education for citizenship: A portrait of current practice in Scottish schools and pre-school centres*. HM Inspectorate of Education.

HMIE. (2006b). *Improving Scottish education*. HM Inspectorate of Education.

HMIE. (2007). *Modern studies - A portrait of current practice in Scottish secondary schools*. HM Inspectorate of Education.

Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.

Hooghe, M., Dassonneville, R., & Marien, S. (2015). The impact of education on the development of political trust: Results from a five-year panel study among late adolescents and young adults in Belgium. *Political Studies, 63*, 123-141.

Hooghe, M., & Wilkenfeld, B. (2008). The stability of political attitudes and behaviors across adolescence and early adulthood: A comparison of survey data on adolescents and young adults in eight countries. *Journal of Youth Adolescence, 37*, 155–167. http://dx.doi.org/10.1007/s10964-007-9199-x.

Hoskins, B., Villalba, C. M., & Saisana, M. (2012). *The 2011 Civic Competence Composite Indicator (CCCI-2).* Luxembourg: Publications Office of the European Union.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. NJ: Erlbaum: Mahwah.

Huisman, J., Pijls, T., Van Hoeij, J., Van Voorst van Beest, K., Boonaerts, Y., & Lens, M. (2003). *Portfolio sociale competenties: Primair onderwijs, vmbo, mbo [Portfolio social competences: Primary-, secondary-. vocational education]*. 's Hertogenbosch, Netherlands: CINOP.

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). *Assessment in game-based learning. Foundations, innovations, and perspective*. New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-3546-4.

Inspectie van het Onderwijs. (2006). *Toezicht op burgerschap en integratie [Supervision of citizenship and integration]*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2010). *De staat van het onderwijs. Onderwijsverslag 2008/2009 [The state of education. Education report 2008/2009]*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2011). *De staat van het onderwijs. Onderwijsverslag 2009/2010 [The state of education. Education report 2009/2010]*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2012). *De staat van het onderwijs. Onderwijsverslag 2010/2011 [The state of education. Education report 2010/2011]*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2013). *De staat van het onderwijs. Onderwijsverslag 2011/2012 [The state of education. Education report 2011/2012]*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2016). *Burgerschap op school: Een beschrijving van burgerschapsonderwijs en de maatschappelijke stage [Citizenship at school: A description of citizenship education and service learning]*. Utrecht, the Netherlands: Inspectie van het Onderwijs.

Jacobson, W., Sleicher, D., & Maureen, B. (1999). Portfolio assessment of intercultural competence. *International Journal of Intercultural Relations, 23*, 467–492. http://dx.doi.org/10.1016/S0147-1767(99)00006-1.

Jeffries, C., & Maeder, D. (2011). Comparing vignette instruction and assessment tasks to classroom observations and reflections. *The Teacher Educator, 46*, 161– 175. http://dx.doi.org/10.1080/08878730.2011.552667.

Jerome, L. (2008). Assessing citizenship education. In J. Arthur, I. Davies, & C. Hahn (Eds.), *The SAGE handbook of education for citizenship and democracy* (pp. 545–559). London, England: SAGE. http://dx.doi.org/10.4135/9781849200486.

Johnson, R.S., Mims-Cox, J.S., & Doyle-Nichols, A. (2010). *Developing portfolios in education: a guide to reflection, inquiry, and assessment*. Los Angeles, CA: Sage.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144.

Karssen, M., van der Veen, I., & Volman, M. (2017). Diversity among bi-ethnic students and differences in educational outcomes and social functioning. *Social Psychology of Education*. Advance online publication. http://dx.doi.org/10.1007/s11218-017-9394-x.

Karsten, S. (2016). *De hoofdstroom in de Nederlandse onderwijsdelta: Een nuchtere balans van het mbo [The mainstream in the Dutch education delta: A sober balance of vocational education]*. Apeldoorn, Netherlands: Maklu.

Karsten, S., Visscher, A. J., Dijkstra, A. B., & Veenstra, R. (2010). Towards standards for the publication of performance indicators in the public sector: The case of schools. *Public Administration, 88*, 90–112. http://dx.doi.org/10.1111/j.1467-9299.2010.01812.x.

Keating, A., Benton, T., & Kerr, D. (2011). Tracing the trends and transitions in young people's citizenship practices: What are the implications for researching citizenship and citizenship education? *Educational Research, 53*, 223–235. http://dx.doi.org/10.1080/00131881.2011.572371.

Keating, A., Kerr, D., Benton, T., Mundy, E., & Lopes, J. (2010). *Citizenship education in England 2001–2010: young people's practices and prospects*

for the future: The eighth and final report from the citizenship education longitudinal study (CELS). London, England: Department for Education.

Kerr, D., Keating, A., & Ireland, E. (2009). *Pupil assessment in citizenship education: Purposes, practices and possibilities. Report of a CIDREE collaborative project*. Slough, England: NFER/CIDREE.

Kerr, D., Smith, A., & Twine, C. (2008). Citizenship education in the United Kingdom. In J. Arthur, I. Davies, & C. Hahn (Eds.), *The SAGE handbook of education for citizenship and democracy* (pp. 252-263). London: SAGE Publications.

Kerr, D., Sturman, L., Schulz, W., & Burge, B. (2010). *ICCS 2009 European report. Civic knowledge, attitudes, and engagement among lower-secondary students in 24 European countries*. Amsterdam: IEA.

Kisby, B., & Sloam, J. (2012). Citizenship, democracy and education in the UK: Towards a common framework for citizenship lessons in the four home nations. *Parliamentary Affairs, 65*, 68-89.

Klenowski, V. (2002). *Developing portfolios for learning and assessment*. London, England: RoutledgeFalmer.

Knight Abowitz, K., & Harnish, J. (2006). Contemporary discourses of citizenship. *Review of Educational Research, 76*, 653–690. http://dx.doi.org/10.3102/00346543076004653.

Kohlberg, L. (1981). *Essays on moral development, vol. I: The philosophy of moral development.* San Francisco, CA: Harper & Row.

Kostons, D., van Gog, T. & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*(2), 121–132. http://dx.doi.org/10.1016/j.learninstruc.2011.08.004.

Landis, J. R., & Koch, K. B. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lauglo, J., & Oia, T. (2006). *Education and civic engagement among Norwegian youths*. Oslo: NOVA.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815355.

Lawy, R., & Biesta, G. (2006). Citizenship-as-practice: The education implications of an inclusive and relational understanding of citizenship. *British Journal of Educational Studies, 54*(1), 34-50.

Ledoux, G., Meijer, J., Van der Veen, I., & Breetvelt, I. (2013). *Meetinstrumenten voor sociale competenties, metacognitie en advanced skills: Een*

inventarisatie [*Assessment instruments for social competences, metacognition and advanced skills: An inventory*]. Amsterdam, Netherlands: Kohnstamm Instituut.

Leeman, Y., & Pels, T. (2006). Citizenship Education in the Dutch Multi-ethnic Context. *European Education, 38*(2), 64-75.

Leenders, H., Veugelers, W., & De Kat, E. (2008). Teachers' views on citizenship education in secondary education in the Netherlands. *Cambridge Journal of Education, 38*(2), 155–170. http://dx.doi.org/10.1080/03057640802063106.

Liden, H., & Odegard, G. (2002). *Ungt engasjement. Ungdoms engasjement og lokalpolitiske deltagelse [Young engagement. Youth engagement and local political participation]*. Oslo: NOVA.

Lindblad, S., Lundahl, L., Lindgren, J., & Zackari, G. (2002). Educating for the New Sweden? *Scandinavian Journal of Educational Research, 46*(3), 283-303.

Listhaug, O., & Gronflaten, L. (2007). Civic decline? Trends in political involvement and participation in Norway, 1965-2001. *Scandinavian Political Studies, 30*(2), 272-299.

Ljunggren, C., & Ost, I. U. (2010). Skolors och larares kontrovershantering [Schools and teachers controversy handling]. In Skolverket, *Skolor som politiska arenor [The school as political arena]* (pp. 18-43). Stockholm: Fritzes kundservice.

LTS. (2000). *Education for citizenship in Scotland: A paper for discussion and consultation*. Dundee: Learning and Teaching Scotland.

LTS. (2002). *Education for citizenship in Scotland: A paper for discussion and development*. Dundee: Learning & Teaching Scotland.

LTS. (2007). *Participation and learning*. Learning and Teaching Scotland.

Lundahl, L. (2002). From centralisation to decentralisation: Governance of education in Sweden. *European Educational Research Journal, 1*(4), 625-636.

Lundahl, L. (2005). A matter of self-governance and control. *European Education, 37*(1), 10-25.

Lundahl, L., Arreman, I. E., Lundstrom, U., & Ronnberg, L. (2010). Setting things right? Swedish upper secondary school reform in a 40-year perspective. *European Journal of Education, 45*(1), 46-59.

Maitles, H. (2000). *Political education in Scotland: An assessment of effectiveness*. Paper presented at the European Conference on Education Research. Edinburgh.

Maitles, H., & Deuchar, R. (2007). 'Why do they never listen to us!' Participation and democratic practice in schools. In A. Ross, *Citizenship Education in Society* (pp. 71-84). London: CiCe.

Maslowski, R., van der Werf, G., Oonk, H., Naayer, H., & Isac, M. M. (2012). *Burgerschapscompetenties van leerlingen in de onderbouw van het voortgezet onderwijs: Eindrapport van de International Civic and Citizenship Education Study in Nederland [Citizenship competences of pupils in lower-secondary education: Final report of the ICCS in the Netherlands]*. GION, Rijksuniversiteit Groningen.

McCrone, D. (2003). Culture, nationalism and Scottish education: Homogeneity and diversity. In T. Bryce, & W. Humes (Eds.), *Scottish education* (pp. 239-249). Edinburgh: Edinburgh University Press.

McDonald, B. (2012). Portfolio assessment: Direct from the classroom. Assessment & *Evaluation in Higher Education, 37*, 335–347. http://dx.doi.org/10.1080/02602938.2010.534763.

Meier, S. L., Rich, B. S., & Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: factors related to discrepancies in scoring. *Assessment in Education: Principles, Policy & Practice, 13*(1), 69-95. http://dx.doi.org/10.1080/09695940600563512.

Mertler, C. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation, 7*(25). http://pareonline.net/getvn.asp?v=7&n=25.

Mikkelsen, R. (2004). Northern Europe regional synthesis. In Council of Europe, *All-European Study on Education for Democratic Citizenship Policies* (pp. 83-90). Strasbourg: Council of Europe Publishing.

Mikkelsen, R., & Fjeldstad, D. (2013). Norway. In J. Ainley, W. Schulz, & T. Friedman (Eds.), *ICCS 2009 Encyclopedia. Approaches to civic and citizenship education around the world* (pp. 313-321). Amsterdam: International Association for the Evaluation of Educational Achievement.

Mikkelsen, R., Buk-Berge, E., Ellingsen, H., Fjeldstad, D., & Sund, A. (2002). *Demokratisk beredskap og engasjement hos 9.-klassinger i Norge og 27 andre land [Democratic readiness and commitment of 9th-graders in Norway and 27 other countries]*. Oslo: Institutt for larerutdanning og skoleutvikling, Universitetet i Oslo.

Mikkelsen, R., Fjeldstad, D., & Lauglo, J. (2011). *Morgendagens samfunnsborgere. Norske ungdomsskoleelevers prestasjoner og svar på spørsmål i den internasjonale demokratiundersøkelsen ICCS [Tomorrow's*

citizens. Norwegian secondary school students' performance and answers in the international democracy survey ICCS]. Oslo: ILS og forfatterne.

Mills, I. (2004). Citizenship: Pupil involvement in Scottish schools. *Pedagogy, Culture and Society, 12*(2), 259-277.

Ministerie van Onderwijs, Cultuur en Wetenschap. (2005). *Memorie van Toelichting bij de Wijziging van de Wet op het primair onderwijs, de Wet op de expertisecentra en de Wet op het voortgezet onderwijs in verband met de bevordering van actief burgerschap en sociale integratie*. 's-Gravenhage: Sdu Uitgevers.

Ministerie van Onderwijs, Cultuur en Wetenschap. (2006a). *Kerndoelen primair onderwijs [Core objectives for primary education]*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap..

Ministerie van Onderwijs, Cultuur en Wetenschap. (2006b). *Kerndoelen onderbouw voortgezet onderwijs [Core objectives for lower-secondary education]*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap..

Ministerie van Onderwijs, Cultuur en Wetenschap. (2013). *Burgerschap in het onderwijs [Citizenship in education]*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.

Ministerie van Onderwijs, Cultuur & Wetenschap. (2017). *Dashboard sectorakkoord vo 2017*.

Mislevy, R., Oranje, A., Bauer, M., Von Davier, A., Hao, J., Corrigan, S., & John, M. (2014). *Psychometric considerations in game-based assessment*. Princeton, NJ: ETS. http://www.instituteofplay.org/wp-content/uploads/2014/02/GlassLab_GBA1_WhitePaperFull.pdf.

Munn, P., & Arnott, M. (2009). Citizenship in Scottish schools: The evolution of education for citizenship from the late twentieth century to the present. *History of Education, 38*(3), 437-454.

Munn, P., Clark, G., Dargie, R., Meldrum, G., & Ross, H. (2004a). *Education for citizenship: Mapping the social subjects curriculum*. Scottish Executive.

Munn, P., Stead, J., McLeod, G., Brown, J., Cowie, M., McCluskey, G., et al. (2004b). Schools for the 21st century: The national debate on education in Scotland. *Research Papers in Education, 19*(4), 433-452.

Munniksma, A., Dijkstra, A.B., Van der Veen, I., Ledoux, G., Van de Werfhorst, H., & Ten Dam, G. (2017). *Burgerschap in het voortgezet onderwijs. Nederland in vergelijkend perspectief [Citizenship in secondary education. The Netherlands in comparative perspective]*. Amsterdam: Amsterdam University Press.

Nie, N., Junn, J., & Stehlik-Barry, K. (1996). *Education and democratic citizenship in America*. Chicago, IL: University of Chicago Press.

Nieuwelink, H. (2016). *Becoming a democratic citizen*.

Nieuwelink, H., Dekker, P., Geijsel, F., & Ten Dam, G. (2016). 'Democracy always comes first': adolescents' views on decision-making in everyday life and political democracy. *Journal of Youth Studies, 19*(7), 990-1006. http://dx.doi.org/10.1080/13676261.2015.1136053.

Norwegian Directorate for Education and Training. (2006a). *Core curriculum for primary, secondary and adult education in Norway*. Oslo, Norway: Directorate for Education and Training (Udir).

Norwegian Directorate for Education and Training. (2006b). *Curriculum for religion, philosophies of life and ethics*. Oslo: Directorate for Education and Training.

Norwegian Directorate for Education and Training. (2006c). *The quality framework*. Oslo: Directorate for Education and Training.

Norwegian Directorate for Education and Training. (2008). *Equity in Education for all – understanding central concepts*. Oslo: Norwegian Directorate for Education and Training.

Norwegian Directorate for Education and Training. (2013). *Curriculum for social studies (SAF1-03) [Laereplan i samfunnsfag]*. Oslo: Directorate for Education and Training.

OECD. (2004). *Country note Norway: Equity in education thematic review*. Paris: OECD.

OECD. (2007). *OECD review of the quality and equity of education outcomes in Scotland. Diagnostic report*. Scottish Executive.

OECD. (2012). *Education at a glance 2012: OECD indicators*. Paris: OECD.

OECD. (2013a). *Country note Norway: Results from PISA 2012*. Paris: OECD.

OECD. (2013b). *Education policy outlook: Norway*. Paris: OECD.

Olson, M. (2012). Theme: Citizenship education under liberal democracy. *Utbildning & Demokrati, 21*, 17–27.

Onderwijsraad. (2003). *Onderwijs en burgerschap [Education and citizenship]*. Den Haag: Artoos.

Onderwijsraad. (2012). *Verder met burgerschap in het onderwijs [Taking citizenship further in education]*. Den Haag: Onderwijsraad.

Opheim, V. (2004). *Equity in education. Country analytical report on Norway*. Oslo: NIFU STEP.

Ormston, R., Curtice, J., McConville, S., & Reid, S. (2011). *Scottish Social Attitudes survey 2010: Attitudes to discrimination and positive action*. Scottish Centre for Social Research.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative purposes revisited: A review. *Educational Research Review, 9*, 129-144. http://dx.doi.org/10.1016/j.edurev.2013.01.002.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice, 21*(2), 133-148. http://dx.doi.org/10.1080/0969594X.2013.877872.

Paulson, L., Paulson, P., & Meyer, C. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48*, 60–63.

Peschar, J. L., Hooghoff, H., Dijkstra, A. B., & Ten Dam, G. (2010). Een agenda voor de toekomst: op weg naar een kennisbasis voor burgerschapsonderwijs [An agenda for the future: On the road to a knowledge base for citizenship education]. In J. L. Peschar, H. Hooghoff, A. B. Dijkstra, & G. Ten Dam (Eds.), *Scholen voor burgerschap. Naar een kennisbasis voor burgerschapsonderwijs [Schooling for citizenship. To a knowledge base for citizenship education]*. Antwerpen-Apeldoorn: Garant.

Pike, M. (2007). Values and visibility: The implementation and assessment of citizenship education in schools. *Educational Review, 59*, 215–229. http://dx.doi.org/10.1080/00131910701255020.

Potter, J. (2006). *Education for values and citizenship: Five nations conference Edinburgh, 2006*. Edinburgh: Institute for Global Ethics UK Trust.

Quintelier, E., & Van Deth, J. (2014). Supporting democracy: Political participation and political attitudes. Exploring causality using panel data. *Political Studies, 62*, 153–171. http://dx.doi.org/10.1111/1467-9248.12097.

Priestley, M., & Humes, W. (2010). The development of Scotland's Curriculum for Excellence: Amnesia and deja vu. *Oxford Review of Education, 36*(3), 345-361.

Razaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing, 15*(1), 18-39. http://dx.doi.org/10.1016/j.asw.2010.01.003.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435–448. http://dx.doi.org/10.1080/02602930902862859.

Redecker, C. (2013). *The use of ICT for the assessment of key competences*. Luxembourg, Luxembourg: Publications Office of the European Union.

Richardson, M. (2010). Assessing the assessment of citizenship. *Research Papers in Education, 25*, 457–478. http://dx.doi.org/10.1080/02671520903045566.

Roelofs, E., & Sluijter, C. (2016). Meetinstrumenten [Measurement instruments]. In Skills Platform, *Skills voor de toekomst: Een onderzoeksagenda [Skills for the future: A researchagenda]* (pp.63-72). Den Haag: Ministerie van Onderwijs, Cultuur & Wetenschap.

Ross, H., Munn, P., & Brown, J. (2007). What counts as student voice in active citizenship case studies? Education for citizenship in Scotland. *Education, Citizenship and Social Justice, 2*(3), 237-256.

Russo, S., & Stattin, H. (2017). Stability and change in youths' political interest. *Social Indicators Research, 132*(2), 643-658. http://dx.doi.org/10.1007/s11205-016-1302-9.

Rychen, D. S., & Salganik, L. H. (2003). *Key Competencies for a Successful Life and Well-functioning Society*. Göttingen, Germany: Hogrefe & Huber Publishers.

Rye, S. A., & Rye, J. F. (2011). Ungdom, internett og samfunnsengasjement [Youth, the internet and civic engagement]. *Tidsskrift for Ungdomsforskning [Journal of Youth Research], 11*(2), 25-47.

Samuelsson, M. (2013). Deliberativ demokrati i den norska skolan. Ger larare uttryck for deliberative uppfattningar nar det kommer till demokrati och demokratiutbildning? [Deliberative democracy in Norwegian schools]. *Utbildning & Demokrati [Education & Democracy], 22*(1), 47-63.

Sandstrom Kjellin, M., & Stier, J. (2008). Citizenship in the classroom: Transferring and transforming transcultural values. *Intercultural Education, 19*(1), 41-51.

Schuitema, J., Ten Dam, G., & Veugelers, W. (2008). Teaching strategies for moral education: A review. *Journal of Curriculum Studies, 40*, 69–89. http://dx.doi.org/10.1080/00220270701294210.

Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam, Netherlands: IEA.

Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., & Friedman, T. (2017). *Becoming citizens in a changing world: ICCS 2016 international report*. Amsterdam, The Netherlands: IEA.

Schulz, W., Fraillon, J., & Ainley, J. (2013). Measuring young people's understanding of civics and citizenship in a cross-national study. *Educational Psychology: An International Journal of Experimental*

Educational Psychology, 33, 334–356.
http://dx.doi.org/10.1080/01443410.2013.772776.

Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). *International civic and citizenship education study: Assessment framework.* Amsterdam, the Netherlands: IEA.

Scottish Executive. (2002). *School boards in Scottish schools*. Scottish Executive National Statistics.

Scottish Executive. (2006). *Scottish Schools (Parental Involvement) Act 2006*. Scottish Executive.

Scottish Executive. (2009). *Assessment for Curriculum for Excellence: Strategic vision and key principles*. Scottish Executive.

Scottish Government. (2013). *Statistical bulletin education series: Summary statistics for schools in Scotland*. Edinburgh: Scottish Government.

Segers, M., Gijbels, D., & Thurlings, M. (2008). The relationship between students' perceptions of portfolio assessment practice and their approaches to learning. *Educational Studies, 34*, 35–44. http://dx.doi.org/10.1080/03055690701785269.

Skolverket. (2001). *Attityder till skolan 2000 [Attitudes towards school]*. Stockholm: Liber.

Skolverket. (2003). *Ung i demokratin [Young people in democracy]*. Stockholm: Skolverket.

Skolverket. (2004). *Nationella utvärderingen av grundskolan 2003. Sammanfattande huvudrapport. [National evaluation of compulsory school 2003. Summary of the main report]*. Stockholm: Skolverket.

Skolverket. (2006). *Individ- och klassvariation i grundskolan åk. 9 [Individual and class variation in compulsory school year 9]*. Stockholm: Skolverket.

Skolverket. (2009a). *Compulsory school syllabuses 2000. Revised version 2008*. Stockholm: Skolverket.

Skolverket. (2009b). *What influences Educational Achievement in Swedish Schools? A Systematic Review and Summary Analysis*. Stockholm: Skolverket.

Skolverket. (2011a). *Curriculum for the compulsory school, preschool class and the leisure-time centre 2011*. Stockholm: Skolverket.

Skolverket. (2011b). *Morgondagens medborgare. ICCS 2009: Svenska 14-åringars kunskaper, värderingar och deltagande i internationell belysning [Tomorrow's citizens. ICCS 2009: Swedish 14-year-olds' values and participation in international perspective]*. Stockholm: Skolverket.

Skolverket. (2011c). *Skolan och medborgarskapandet [Schooling for citizenship]*. Stockholm: Fritzes kundservice.

Skolverket. (2013). *Resource Allocation to Schools by Municipalities. A Summary of Two Reports from the Swedish National Agency for Education*. Stockholm: Skolverket.

Solhaug, T. (2003). *Utdanning til demokratisk medborgerskap [Education for democratic citizenship]*. Oslo: Institutt for larerutdanning og skoleutvikling, Universitetet i Oslo.

Solost, M. (2011). *"Norske elever viser gode kunnskaper om demokrati" Men hva med helhetlig demokratiske kompetanse? ["Norwegian pupils show good knowledge of democracy" But what about their overall democratic competence?]*. Oslo: Institutt for larerutdanning og skoleutvikling, Universitetet i Oslo.

SOU. (2002). *Skollag för kvalitet och likvärdighet [Schools for quality and equality]*. Stockholm: Ministry of Education.

Squire, K. (2006). Content to context: Videogames as designed experience. *Educational Researcher, 35*, 19–29. http://dx.doi.org/10.3102/0013189X035008019.

Squire, K., & Jenkins, H. (2003). Harnessing the power of games in education. *Insight, 3*, 5–33.

Starkey, H. (2005). Language teaching for democratic citizenship. In A. Osler & H. Starkey (Eds.), *Citizenship and Language Learning: International Perspectives* (pp. 23–40). Staffordshire: Trentham Book Limited

Stolz, P. (2011). Canons and communities: Children and social cohesion in Sweden and the Netherlands. *Education, Citizenship and Social Justice, 6(*1), 39-52.

Stray, J. H. (2009). *Demokratisk medborgerskap i norsk skole? En kritisk analyse [Democratic citizenship in Norwegian schools? A critical analysis]*. Oslo: Det utdanningsvitenskapelige fakultet, Universitetet i Oslo Unipub.

Susi, T., Johannesson, M., & Backlund, P. (2007). *Serious games – an overview technical report HS-IKI-TR-07-001*. Sweden: University of Skövde.

Taras, M. (2005). Summative and formative: Some theoretical reflections. *British Journal of Educational Studies, 53,* 466–478. http://dx.doi.org/10.1111/j.1467-8527.2005.00307.x.

Ten Dam, G., Geijsel, F., Ledoux, G., & Meijer, J. (2013). Citizenship of students and social desirability: Living apart together? *International Journal of Educational Research, 62*, 229–238. http://dx.doi.org/10.1016/j.ijer.2013.10.001.

Ten Dam, G., Geijsel, F., Reumerman, R., & Ledoux, G. (2011). Measuring young people's citizenship competences. *European Journal of Education, 46*, 354–478. http://dx.doi.org/10.1111/j.1465-3435.2011.01485.x.

Ten Dam, G., & Volman, M. (2003). A life jacket or an art of living: Inequality in social competence education. *Curriculum Inquiry, 33*(2), 117-137.

Ten Dam, G., & Volman, M. (2004). Critical thinking as a citizenship competence: teaching strategies. *Learning and Instruction, 14*, 359-379. http://dx.doi.org/10.1016/j.learninstruc.2004.01.005.

Ten Dam, G., & Volman, M. (2007). Educating for adulthood or for citizenship: social competence as an educational goal. *European Journal of Education, 42*(2), 281–298. http://dx.doi.org/10.1111/j.1465-3435.2007.00295.x.

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9(2). http://pareonline.net/getvn.asp?v=9&n=2.

Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning – A review of research studies. *Studies in Educational Evaluation, 37*, 25–34. http://dx.doi.org/10.1016/j.stueduc.2011.03.004.

Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen.* Amsterdam: IEA.

Turper, S., Iyengar, S., Aarts, K., & Van Gerven, M. (2015). Who is less welcome? The impact of individuating cues on attitudes towards immigrants. *Journal of Ethnic and Migration Studies, 41*, 239–259. http://dx.doi.org/10.1080/1369183X.2014.912941.

Tweede Kamer. (2018). *Handelingen 28 maart 2018 dertigledendebat over burgerschapsonderwijs en maatschappijleer [Minutes 28 March 2018, debate on citizenship education and civic studies]*.

Utdanningsdirektoratet. (2012). *Kartlegging av temaer knyttet til menneskerettigheter og demokratiforståelse i læreplanverket [Survey of topics relating to human rights and democratic understanding in the curriculum]*. Oslo: Utdanningsdirektoratet.

Van den Bos, P., Burghout, C., & Joosten-ten Brink, D. (2014). Toetsen met rubrics [Assessment through rubrics]. In: H. Van Berkel, A. Bax, D. Joosten-ten Brinke (eds.). *Toetsen in het hoger onderwijs [Testing in higher education]* (pp. 193-204). Houten, Netherlands: Bohn Stafleu van Lochem. http://dx.doi.org/10.1007/978-90-368-0239-0_1.

Van Gunsteren, H. (1998). *A theory of citizenship: Organizing plurality in contemporary democracies*. Boulder, CO: Westview Press.

Vandercruysse, S., Vandewaetere, M., & Clarebout, G. (2012). Game based learning: A review on the effectiveness of educational games. In M. Cruz-Cunha (Ed.), *Handbook of research on serious games as educational, business, and research tools* (pp. 628–647). Hershey, PA: IGI Global. http://dx.doi.org/10.4018/978-1-4666-0149-9.ch032.

Veugelers, W. (2007). Creating critical-democratic citizenship education: empowering humanity and democracy in Dutch education. *Compare, 37*(1), 105-119.

Wagenaar, H., Keune, K., & Van Weerden, J. (2012). *Balans oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie [Balance judging and reasoning as part of active citizenship and social integration]*. Arnhem, Netherlands: Cito.

Wagenaar, H., van der Schoot, F., & Hemker, B. (2011). *Balans actief burgerschap en sociale integratie. Uitkomsten van de peiling 2009 [Balance active citizenship and social integration. Outcomes of the 2009 survey]*. Arnhem: Cito.

Welzel, C. (2007). Are levels of democracy affected by mass attitudes? Testing attainment and sustainment effects on democracy. *International Political Science Review, 28*(4), 397-424.

Westheimer, J., & Kahne, J. (2004). What kind of citizen? The politics of educating for democracy. *American Educational Research Journal, 41*(2), 1–26. http://dx.doi.org/10.3102/00028312041002237.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*, 3–14. http://dx.doi.org/10.1016/j.stueduc.2011.03.001.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*, 537–548. http://dx.doi.org/10.1080/0141192960220502.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van der Spek, E. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*, 249–265. http://dx.doi.org/10.1037/a0031311.

Zwaans, A., Ten Dam, G., & Volman, M. (2006). Teachers' goals regarding social competence. *European Journal of Teacher Education, 29*(2), 181-202.

CONTRIBUTIONS OF AUTHORS

Chapter 2 is based on:

Daas, R. (2014). Linking citizenship education policy to students' citizenship competences in the Netherlands, Norway, Scotland and Sweden. In A. B. Dijkstra, & P. I. de la Motte (Eds.), *Social outcomes of education: The assessment of social outcomes and school improvement through school inspections* (pp. 73–99). Amsterdam, Netherlands: Amsterdam University Press.

Contributions:

Remmert Daas prepared the study, conducted the analyses, and wrote the manuscript. Empirical data on students' citizenship competences originated from the International Civic and Citizenship Education Study 2009. While conducting the study Remmert Daas was supervised by Sjoerd Karsten, who provided regular feedback on draft versions of the manuscript. Anne Bert Dijkstra as editor of the book in which the study was published provided feedback on the final draft.

Chapter 3 is based on:

Daas, R., Dijkstra, A. B., Karsten, S., & Ten Dam, G. (submitted). Assessing citizenship competences using a rubric-based approach.

Contributions:

Remmert Daas reviewed the literature, held panel discussions with teachers, students and researchers, collaborated with teachers to conduct the assessment, assessed students' materials, trained assessors to also assess students' materials, conducted the analyses, reported results back to teachers, and drafted various versions of the manuscript. Geert ten Dam and Anne Bert Dijkstra contributed to the conceptual framework and design of the rubrics and supervised Remmert Daas while developing the rubrics. Anne Bert Dijkstra and Sjoerd Karsten supervised Remmert Daas while writing the manuscript and audited the analysis and interpretation of the data. They discussed the various steps of the research and contributed to reviews and revisions of the manuscript.

Chapter 4 is based on:

Daas, R., Dijkstra, A. B., & Karsten, S. (submitted). Assessing citizenship attitudes using rubrics.

Contributions:

Remmert Daas reviewed the literature, analyzed the data and drafted the various versions of the manuscript. The data was collected as part of the study reported on in Chapter 3. The research team for this article further consisted of Anne Bert

Dijkstra and Sjoerd Karsten, who together supervised Remmert Daas while writing the manuscript and audited the analysis and interpretation of the data. They discussed the various steps of the research and contributed to reviews and revisions of the manuscript.

Chapter 5 is based on:

Daas, R., Ten Dam, G., & Dijkstra, A. B. (2016). Contemplating modes of assessing citizenship competences, *Studies in Educational Evaluation, 51*, 88-95. http://dx.doi.org/10.1016/j.stueduc.2016.10.003.

Contributions:

Remmert Daas reviewed the literature and drafted the various versions of the manuscript. The research team for this article further consisted of Geert ten Dam and Anne Bert Dijkstra, who together supervised Remmert Daas. The research team discussed the various steps of the research. The team collaboratively conceptualized and designed the study, deliberated about the theoretical framework, the results and the conclusions. The supervisors contributed to the conceptual framework and analyses, and reviewed and revised the manuscript.

NEDERLANDSTALIGE SAMENVATTING

Dit proefschrift is gericht op het beoordelen van burgerschapscompetenties. In de vier studies die we uitvoerden wordt dit vraagstuk vanuit verschillende invalshoeken uitgewerkt. In dit laatste hoofdstuk worden de belangrijkste bevindingen gepresenteerd. Hieronder volgt eerst een samenvatting van de vier deelstudies. Vervolgens bespreken we de bijdragen die deze studies leveren aan ons begrip van het beoordelen van burgerschapscompetenties, en welke beperkingen daarbij in ogenschouw genomen moeten worden. We sluiten af met de implicaties van onze bevindingen voor toekomstige metingen van burgerschapscompetenties van leerlingen.

Belangrijkste bevindingen

Burgerschapscompetenties van leerlingen reflecteren landelijk onderwijsbeleid

De resultaten van leerlingen op metingen van burgerschapscompetenties kunnen geaggregeerd worden naar klas-, school- en zelfs nationaal niveau. Omdat de *International Civics and Citizenship Education Study* (ICCS) een nationaal-representatieve steekproef gebruikt, kunnen de resultaten van leerlingen tussen landen vergeleken worden (Schulz et al., 2010). Met behulp van deze resultaten kunnen we evalueren of de resultaten van leerlingen ook landelijk beleid en discours rond burgerschapsonderwijs reflecteren. De eerste deelstudie van dit proefschrift beschrijft de doelen en uitkomsten van burgerschapsonderwijs in Nederland, Noorwegen, Schotland en Zweden. Deze landen zijn in verschillende opzichten vergelijkbaar, maar laten elk een eigen benadering van burgerschapsonderwijs zien. Op basis van analyse van beleidsnotities, onderzoeksrapporten, inspectierapporten en wetenschappelijke publicaties wordt voor elk van deze vier landen een beeld geschetst van het burgerschapsonderwijs. Deze beelden worden vergeleken met de resultaten van leerlingen op vijf schalen uit ICCS 2009: burgerschapskennis, houding tegenover basisprincipes van een democratie, houding tegenover gelijke rechten voor alle etnische groepen, verwachting van jongeren om later gebruik te gaan maken van hun stemrecht, en deelname aan burgerschapsactiviteiten op school. In alle vier de landen is de aandacht voor burgerschapsonderwijs de afgelopen jaren toegenomen. Er zijn echter verschillen in de mate waarin bepaalde thema's besproken worden. Omdat data uit ICCS 2009 maar één meetmoment omvatten is het doel niet causale uitspraken te doen, maar om samenhang tussen beleid en uitkomsten op

leerlingenniveau te onderzoeken. Omdat Schotland niet deel nam aan ICCS 2009 is een vergelijking met de resultaten van Schotse leerlingen niet mogelijk.

De aandacht voor burgerschapsonderwijs in Nederland lijkt in sterke mate ingegeven door noties over de integratie van immigranten. Beleid rond burgerschapsonderwijs in Nederland benadrukt het belang van 'samenleven'. Verschillende studies laten echter zien dat er weinig ontwikkelingen plaatsvinden in het onderwijsaanbod hiertoe. Leerlingen in Nederland scoren vergelijkbaar aan het internationaal gemiddelde op burgerschapskennis, en onder het internationaal gemiddelde op de andere vier geselecteerde aspecten van burgerschapscompetenties. De resultaten laten ook zien dat verschillen tussen scholen in Nederland groter zijn dan in Noorwegen of Zweden; waarschijnlijk vanwege de differentiatie in het onderwijsstelsel naar verschillende onderwijsniveaus.

Burgerschapsonderwijs in Noorwegen lijkt in vergelijking met de andere drie landen weinig aandacht te krijgen in onderwijsbeleid. De resultaten van leerlingen op metingen van burgerschapscompetenties zijn in het verleden hoog geweest, en onderwijsbeleid lijkt zich vooral te richten op cognitieve uitkomsten. Het curriculum bevat echter verschillende mechanismen om burgerschapscompetenties te bevorderen, zoals verschillende sociale vakken. Leerlingen in Noorwegen scoren boven het internationale gemiddelde op alle vijf de geselecteerde aspecten van burgerschapscompetenties.

Burgerschapsonderwijs in Schotland lijkt sterk gericht op participatie door leerlingen. Ontwikkeling van burgerschap is een van vijf centrale thema's in het onderwijs, en burgerschapsonderwijs is zowel cross-curriculair als in het vak 'social studies' ingebed. Op basis van deze randvoorwaarden zou verwacht kunnen worden dat leerlingen in Schotland bovengemiddeld scoren op de meeste aspecten van burgerschapscompetenties. Schotland nam echter geen deel aan ICCS, wat kenmerkend lijkt voor de aandacht die zich meer op het onderwijsproces richt dan op onderwijsopbrengsten (vgl. Dijkstra & De la Motte, 2014).

De afgelopen jaren heeft in Zweden, net als in veel andere westerse landen, meer liberalisering plaatsgevonden. Zweden staat echter bekend om een lange traditie van het bevorderen van sociale waarden, en burgerschapsonderwijs lijkt ook in die traditie te passen. Sociale waarden lijken zowel binnen als buiten het onderwijs breed onderschreven te worden. Leerlingen in Zweden scoren tussen hun leeftijdsgenoten in Noorwegen en Nederland in. Ze scoren hoger dan het internationale gemiddelde op burgerschapskennis en houding tegenover gelijke rechten voor alle etnische groepen, gemiddeld op houding tegenover basisprincipes van een democratie en deelname aan burgerschapsactiviteiten op school, en onder

het internationaal gemiddelde op verwachting van jongeren om later gebruik te gaan maken van hun stemrecht.

De analyses laten zien dat metingen van burgerschapscompetenties van leerlingen het discours over (de effectiviteit van) burgerschapsonderwijsbeleid weerspiegelen. Landen hanteren verschillende benaderingen van burgerschapsonderwijsbeleid, en het is relevant hier notie van te nemen bij meten van burgerschapscompetenties van leerlingen (vgl. Hahn, 1999). Simpel gezegd: beleid ten aanzien van burgerschapsonderwijs doet er toe.

Rubrics kunnen worden gebruikt om burgerschapscompetenties te beoordelen
De meeste studies waarin burgerschapscompetenties van leerlingen gemeten worden maken gebruik van enquêtes bestaande uit (meerkeuze) toetsen en vragenlijsten. Er is eerder opgeroepen het beschikbare instrumentaria verder uit te breiden (Jerome, 2008; Kerr, Keating, & Ireland, 2009). De tweede deelstudie is daarom gericht op een alternatieve benadering om burgerschapscompetenties te beoordelen. Hiervoor ontwikkelden we een nieuw beoordelingsinstrument. Omdat bestaande instrumenten vooral meerkeuzetoetsen en vragenlijsten gebruiken, ontwikkelden we een instrument dat een meer open benadering toelaat: rubrics (vgl. Andrade, 2005; Jonsson & Svingy, 2007; Panadero & Jonsson, 2013). We ontwikkelden drie rubrics voor drie 'sociale taken': democratisch handelen, maatschappelijk verantwoord handelen, en omgaan met verschillen (Ten Dam et al., 2011). Deze sociale taken kunnen beschouwd worden als exemplarisch voor hoe jonge mensen in hun dagelijks leven vorm geven aan burgerschap. Deze sociale taken reflecteren ook de aandacht voor 'samenleven' in het Nederlandse burgerschapsonderwijsbeleid (Ten Dam & Volman, 2007). Elke rubric bestaat uit drie componenten: kennis, houding en vaardigheid. Elke component is omschreven op vier niveaus: A t/m D. Elk hoger niveau veronderstelt dat de lagere niveaus ook beheerst worden. De volledige rubrics zijn te vinden in Appendix B. Hieronder wordt een korte beschrijving gegeven van de verschillende niveaus die in de rubrics beschreven worden. In Hoofdstuk 2 en 3 van dit proefschrift wordt een meer gedetailleerde uitwerking en onderbouwing gegeven.

Kennis niveau A beschrijft kennis van democratische principes, zoals anderen geen geweld aandoen. Niveau B beschrijft een begrip van de consequenties van deze principes, zoals beperkingen aan de vrijheid van meningsuiting. Niveau C beschrijft het toepassen van deze kennis, zoals de implicaties van de verzorgingsstaat. Niveau D beschrijft het evalueren van kennis, zoals het evalueren van de oorzaken en gevolgen van politieke beslissingen.

Houding niveau A beschrijft een houding die als legitiem beschouwd kan worden, maar waarbij samenleven als probleemloos beschouwd wordt. Niveau B kan beschouwd worden als een houding waarbij democratische principes impliciet gesteund worden. Niveau C beschrijft een houding waarbij expliciet en bewust steun is voor democratische principes. Niveau D beschrijft een houding waarbij sprake is van sensitiviteit voor complexe situaties, en aandacht voor het bevorderen van democratische houdingen bij anderen.

Vaardigheid niveau A beschrijft het vermogen om het eigen gedrag vorm te geven, zoals vasthouden aan een eigen mening. Niveau B beschrijft het vermogen om het gedrag vorm te geven in relatie tot anderen, zoals aan een discussie kunnen deelnemen. Niveau C beschrijft het vermogen om het eigen gedrag aan te passen in relatie tot anderen en omstandigheden, zoals rekeninghouden met culturele verschillen. Niveau D beschrijft het vermogen om doelgericht te handelen in veeleisende situaties, bijvoorbeeld bij het overtuigen van anderen.

De rubrics die we ontwikkeld hebben zijn bedoeld voor gebruik in het onderwijs. Het ontwikkelen ervan is daarom een iteratief proces geweest. We vertrokken vanuit een conceptuele afweging van welke verschillende beheersingsniveaus onderscheiden zouden kunnen worden op basis van theorie en resultaten uit eerder onderzoek. De eerste versie van deze rubrics hebben we besproken met docenten, leerlingen, docenten in opleiding, docentopleiders en collega onderzoekers. Met name de voorbeelden die in de rubrics beschreven werden zijn op basis van deze gesprekken aanpast.

Uit de gesprekken met leerlingen en docenten kwam naar voren dat de rubrics voor veel leerlingen van 13 of 14 jaar nog moeilijk te begrijpen waren in relatie tot de maatschappij om hen heen. In het vervolg van het onderzoek richtten we ons daarom op leerlingen van 15 jaar en ouder. Leerlingen in havo-4 en eerstejaars mbo niveau 3 en 4 namen deel aan de studie.[1] De burgerschapscompetenties van deze groepen leerlingen zijn niet eerder onderzocht in wetenschappelijk onderzoek, ook al zijn maatschappijleer en burgerschap al geruime tijd onderdeel van het curriculum. We gebruikten de rubrics op twee manieren: 115 studenten maakten een opdracht waarin ze gevraagd werden te reflecteren op een reeks open vragen, op basis waarvan hun burgerschapskennis en -houding met behulp van de rubrics beoordeeld werden; 601 leerlingen

---

[1] In het mbo wordt doorgaans de term 'student' of 'deelnemer' gebruikt. We gebruiken hier de term leerling voor alle deelnemers aan ons onderzoek uit zowel havo als mbo.

beoordeelden hun eigen burgerschapskennis, -houding en -vaardigheden aan de hand van één van de rubrics en gaven daar een toelichting bij.

De antwoorden van leerlingen op de reflectieopdracht werden elk door drie beoordelaars beoordeeld. Op basis van overeenstemming tussen minstens twee beoordelaars konden vrijwel alle leerlingen een niveau van kennis en houding toegekend worden. De inter-beoordelaarsbetrouwbaarheid was redelijk tot goed. De meeste leerlingen scoorden A of B voor zowel kennis als houding, ongeveer tien tot vijftien procent scoorde C, en geen leerlingen scoorden D. De rubrics bleken bruikbaar voor het beoordelen van burgerschapskennis en -houding van leerlingen, maar beoordeling was wel afhankelijk van de kwaliteit van de antwoorden van leerlingen.

Bij het beoordelen van hun eigen burgerschapscompetenties waren vrijwel alle leerlingen in staat een niveau van kennis, houding en vaardigheid te selecteren dat het beste bij hen paste. De meeste leerlingen kozen daarbij voor niveau B of C voor alle drie de componenten. Ruim 90 procent van de leerlingen gaf ook een relevante toelichting bij hun beoordeling. Als we alleen kijken naar leerlingen die ook een adequate toelichting gaven bij het niveau wat zij zichzelf toeschreven scoren de meeste leerlingen niveau B of A. Voor kennis en vaardigheid waren slechts een kwart van de toelichtingen toereikend voor het gekozen niveau; bij houding gold dit voor de helft van de toelichtingen. Het is op basis van de resultaten niet mogelijk om onderscheid te maken tussen leerlingen die zichzelf overschatten, ongemotiveerd waren, of om andere redenen geen toereikende toelichting gaven. Rubrics kunnen op basis van deze resultaten meer geschikt geacht worden voor het beoordelen van burgerschapshoudingen, en leerlingen hebben ondersteuning nodig om hun eigen competenties te beoordelen en – vooral – toe te lichten.

Rubrics verschaffen inzicht in de burgerschapshoudingen van leerlingen

Houdingen vormen een centraal aspect van de ontwikkeling van burgerschapscompetenties. Ze zijn van belang voor het collectief door het onderhouden van democratische samenlevingen waarin democratische waarden en principes in stand gehouden worden (vgl. Dahl, 2000; Eckstein, 1966; Welzel, 2007) en ze zijn van belang voor het individu om effectief te kunnen handelen als sociaal competente burger in een democratisch samenleving (Ten Dam & Volman, 2007).

Burgerschapshoudingen van leerlingen worden doorgaans gemeten door leerlingen een reeks stellingen voor te leggen, waarbij ze op een Likertschaal aangeven in hoeverre zij het met elke stelling eens zijn. Op basis van de tweede

deelstudie vinden we dat rubrics vooral geschikt zijn voor zelfbeoordeling van burgerschapshoudingen. In de derde deelstudie kijken we daarom naar wat zelfbeoordeling van burgerschapshoudingen met behulp van rubrics ons leert over de houdingen van leerlingen ten aanzien van democratisch handelen, maatschappelijk verantwoord handelen, en omgaan met verschillen. De zelfbeoordelingen en toelichtingen van 567 leerlingen worden onderzocht, waarbij tevens wordt gekeken naar verschillen tussen leerlingen op basis van achtergrondkenmerken. Leerlingen waren gemiddeld 16 jaar, en zaten in havo-4 of het eerste jaar van mbo niveau 3 of 4.

Leerlingen kozen het vaakst niveau B voor hun houding ten aanzien van democratisch handelen, en het minst vaak niveau A. Leerlingen lijken impliciet democratische principes en democratisch handelen te steunen. Wanneer we alleen kijken naar leerlingen die een toereikende toelichting bij hun beoordeling geven, scoorden leerlingen het vaakst niveau B of C, en het minst vaak niveau D. Deze leerlingen steunen impliciet of expliciet democratische principes, en – in het geval van niveau C – tonen de intentie deze actief toe te passen in discussies.

Leerlingen scoorden tevens het vaakst niveau B voor hun houding ten aanzien van maatschappelijk verantwoord handelen, en het minst vaak niveau D. Leerlingen lijken de waarde van samenwerking en anderen helpen te steunen. Ze laten het minst vaak een intentie zien om deze houding ook bij anderen te bevorderen. Wanneer we alleen kijken naar leerlingen die een toereikende toelichting bij hun beoordeling geven, scoorden leerlingen nog steeds het vaakst niveau B, en het minst vaak niveau C of D.

Leerlingen scoorden het vaakst niveau C voor hun houding ten aanzien van omgaan met verschillen, en het minst vaak niveau A. Leerlingen lijken sensitiviteit ten aanzien van het omgaan met verschillen van belang te vinden. Wanneer we alleen kijken naar leerlingen die een toereikende toelichting bij hun beoordeling geven scoorden leerlingen nog steeds het vaakst niveau C, en het minst vaak niveau D, wat aangeeft dat ze het minst vaak een intentie laten zien om deze houding ook bij anderen te bevorderen.

De resultaten verschillen enigszins tussen de verschillende sociale taken, maar over het algemeen beoordelen leerlingen hun houding het vaakst niveau B of C. Dit geldt nog steeds wanneer we alleen kijken naar leerlingen die ook een toereikende toelichting geven, al neemt het percentage leerlingen met niveau C of D dan wel af.

De evaluatie van het gebruik van rubrics om burgerschapshoudingen te beoordelen is in dit onderzoek gericht op toepassing in het onderwijs. De zelfbeoordelingen en toelichtingen van leerlingen maken dat het instrument in

essentie relevant is voor het leerproces omdat het gebruikt kan worden om inzicht te krijgen in de persoonlijke invulling die leerlingen geven aan hun houdingen, alsook discrepanties tussen het niveau dat de leerling zichzelf toewijst en de toelichting die daarbij gegeven wordt (vgl. Black & William, 1998; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). In relatie tot het laatste punt moet daarbij wel aangemerkt worden dat de kwaliteit van toelichtingen te wensen overliet, mede gezien maar de helft van de leerlingen een toereikende toelichting gaf.

Door rubrics te gebruiken voor het beoordelen van burgerschapshoudingen ontstaat een conceptueel andere meting van dit concept dan met behulp van items met een Likertschaal. Hogere niveaus van burgerschapshoudingen zijn in de rubrics uitgewerkt als meer diepgang in de afwegingen en sensitiviteit die leerlingen tonen. Uit de resultaten van deze deelstudie komt naar voren dat leerlingen over het algemeen steun laten zien voor democratische principes, anderen helpen, en sensitiviteit tonen voor anderen, maar de meeste leerlingen achten hun houding nog niet gevoelig voor spanningsvolle situaties. Op basis van statische analyses vonden we alleen statistisch kleine verschillen voor leerlingen van verschillende leeftijd, geslacht, onderwijsniveau of migratieachtergrond.

Burgerschapscompetenties kunnen op verschillende manieren beoordeeld worden

Met het beoordelen van burgerschapscompetenties kunnen verschillende doelen beoogd worden. Een bekend onderscheid hierbij is tussen formatief en summatief beoordelen (Broadfoot et al., 2002; William & Black, 1996). Beide typen beoordeling beogen inzichtelijk te maken in hoeverre een leerling een leerdoel beheerst, waarbij formatieve beoordeling ook als doel heeft informatie te genereren om het verdere leerproces te ondersteunen (Taras, 2005). Om zowel formatieve als summatieve beoordeling van burgerschapscompetenties mogelijk te maken is het zinvol te evalueren in hoeverre verschillende instrumenten dit kunnen faciliteren.

In de vierde deelstudie evalueren we vier aanpakken om burgerschapscompetenties te beoordelen: toetsen en vragenlijsten, portfolio's, *game-based assessment*, en vignetten. Op basis van vier concepten formuleren zeven evaluatiecriteria. Gebaseerd op summatieve beoordeling zou een beoordelingsinstrument (1) vergelijking tussen studenten mogelijk moeten maken, en (2) vergelijking van studenten met een norm mogelijk moeten maken. Op basis van formatieve beoordeling zou een beoordelingsinstrument (3) informatie moeten genereren die docenten en leerlingen helpt bij het vormgeven van het leerproces. Op basis van een betekenisvolle beoordeling van burgerschapscompetenties zou

een beoordelingsinstrument (4) een brede uitwerking aan het begrip burgerschap moeten geven zodat leerlingen ruimte hebben eigen invulling en overtuigingen ten aanzien van burgerschap uit te werken, (5) burgerschapscompetenties op verschillende momenten beoordelen om inzicht te geven in de ontwikkeling van leerlingen, en (6) in betekenisvolle context plaatsvinden, waarin leerlingen burgerschap als relevant voor hun eigen situatie ervaren. Ten slotte, omdat beoordeling ook praktisch realiseerbaar moet zijn, zou beoordeling (7) effectief toepasbaar moeten zijn in het onderwijs.

Toetsen en vragenlijsten worden vaak samen gebruikt, en zijn de meest gebruikte instrumenten om burgerschapscompetenties te beoordelen. Door de efficiëntie, objectiviteit en standaardisering zijn ze zeer effectief toepasbaar, met name waar het gaat om summatieve beoordeling. Hoewel resultaten van leerlingen ook gebruikt kunnen worden om onderwijs vorm te geven, geven deze instrument minder persoonlijk relevante informatie; deze instrumenten hebben bijvoorbeeld moeite om onderliggende overtuigingen of afwegingen van leerlingen bloot te leggen. Als gevolg hiervan is het creëren van een betekenisvolle beoordeling moeilijker realiseerbaar met dit type instrumenten.

Portfolio's benaderen beoordeling vanuit een heel andere invalshoek. Leerlingen verzamelen over het algemeen zelf informatie over de eigen competentieontwikkeling. Portfolio's beogen vooral betekenisvolle beoordelingen te realiseren. De potentieel rijke informatie die in portfolio's verzameld wordt, biedt daarbij ook aanknopingspunten voor verdere ontwikkeling van leerlingen. Om het open karakter van portfolio's te bewaren is er echter weinig sprake van standaardisering, waardoor objectieve, summatieve beoordeling gecompliceerd is en waardoor subjectiviteit vaak een rol speelt. Ten slotte zijn portfolio's tijdrovend voor zowel docenten als leerlingen, waardoor effectieve toepasbaarheid in het geding is.

Game-based assessment is een relatief nieuwe benadering van beoordelen in het onderwijs, en voor burgerschapscompetenties in het bijzonder. De complexe omgevingen waarin diverse spelers met elkaar en de omgeving interacteren laten mogelijkheden zien om betekenisvolle situaties te creëren waarin leerlingen rollen op zich kunnen nemen die in het dagelijks leven vaak niet toegankelijk zijn. Omdat alle parameters geregistreerd kunnen worden lijkt vergelijking tussen leerlinggedrag en vergelijking tussen leerlinggedrag en een norm mogelijk. Formatieve beoordeling lijkt minder realiseerbaar, omdat onduidelijk blijft in hoeverre gedrag in een spelsituatie overdraagbaar is naar situaties daarbuiten. De belangrijkste beperking lijkt echter te zijn dat de ontwikkeling van dit type spellen nog veel aandacht vraagt.

Vignetten zijn gedetailleerde omschrijvingen van situaties, en bij instrumenten die daar gebruik van maken worden de reacties van leerlingen gevraagd om inzicht te krijgen in hun overtuigingen en argumentaties. Deze inzichten kunnen gebruikt worden als input voor toekomstig leren en instructie. Door vignetten te standaardiseren kunnen de reacties van leerlingen met elkaar worden vergeleken, en in mindere mate met een norm. Voor een meer betekenisvolle invulling lijkt een uitruil met standaardisering te ontstaan. Vignetten kunnen ook in interviews gebruikt worden, maar vragen dan veel tijd om af te nemen. Naar mate vignetten meer details bevatten kunnen er ook minder in een instrument opgenomen worden vanwege de inspanning die van leerlingen gevraagd wordt. Ten slotte vraagt de ontwikkeling van vignetten veel werk, en deze worden dan ook niet veel gebruikt in het onderwijs.

Alle vier de geëvalueerde instrumenten hebben sterke en zwakke kanten in relatie tot de zeven criteria. Toetsen en vragenlijsten zijn goed afgestemd op summatieve beoordeling, maar hebben moeite een betekenisvolle beoordeling te bieden. Portfolio's bieden een rijkheid aan informatie die betekenisvolle beoordeling kan faciliteren, maar worstelen met betrouwbaarheid en tijdslast. *Game-based assessment* biedt veelbelovende vooruitzichten voor summatieve beoordeling, maar vraagt nog veel ontwikkeling. Vignetten laten een uitruil zien tussen rijke, betekenisvolle contexten en betrouwbare summatieve beoordeling, en vragen veel voorbereiding.

Bijdragen

Op basis van de verschillende deelstudies die in dit proefschrift gepresenteerd worden, worden hier drie meer fundamentele bijdragen van dit proefschrift uitgewerkt. Hierbij gaat het om de diverse functies die beoordelingen van burgerschapscompetenties kunnen vervullen, de waarde van het nieuw ontwikkelde instrument dat burgerschapscompetenties beoordeeld met behulp van rubrics, en de waarde van het omschrijven van niveaus van burgerschapscompetenties.

Functies van beoordelingen
In dit proefschrift bespreken we twee veelgebruikte doelen van beoordelen die in het onderwijs door velen onderscheiden worden: summatief beoordelen heeft algemeen tot doel om prestaties van leerlingen te beoordelen in relatie tot elkaar of een norm; formatief beoordelen heeft tot doel input te leveren voor verder leren. Op basis van deze doelen kan ook beschouwd worden in hoeverre bepaalde instrument meer of minder geschikt zijn voor welk doel. De specifieke functie die

een beoordeling vervult kan echter in meer detail bekeken worden wanneer we kijken naar de manier waarop een instrument ingezet wordt. We bekijken deze functies op drie niveaus: op dat van individuele leerlingen, groepen leerlingen op klas- of schoolniveau, en groepen scholen op onderwijsstelselniveau.

Beoordelen van burgerschapscompetenties van individuele leerlingen kan verschillende functies hebben. Beoordeling kan bijvoorbeeld dienen om een rapport(cijfer) op te baseren, waarin een overzicht wordt gegeven van de inspanningen en prestaties van een leerling. Zowel docenten als leerlingen kunnen deze cijfers gebruiken om leerontwikkeling te monitoren, en daar vervolgacties op te baseren (vgl. Jerome, 2010). Beoordeling kan ook dienen om leerlingen te motiveren, of zelfs duidelijk te maken dat een bepaald thema belangrijk is 'omdat het op de toets komt' (vgl. Richardson, 2010). Als beoordeling plaatsvind vooraf of tijdens het leerproces, kunnen de uitkomsten ook gebruikt worden om leerlingen leerdoelen te laten formuleren.

Docenten kunnen de resultaten van leerlingen ook gebruiken om de effectiviteit van de eigen lessen te evalueren, of resultaten met collega's of eerdere jaren vergelijken. Wanneer gestandaardiseerde instrumenten gebruikt worden kunnen resultaten ook tussen scholen vergeleken worden om de eigen praktijk te evalueren. Andere stakeholders, zoals schoolleiders, bestuurders, ouders, of de onderwijsinspectie, kunnen deze resultaten ook gebruiken als een indicator voor schoolkwaliteit. De resultaten kunnen gebruikt worden om voorgang van leerlingen of scholen te monitoren, (beleids-)beslissingen op te baseren, voor evaluatie, of verantwoording door scholen (Dijkstra & De la Motte, 2014). De eerste deelstudie van dit proefschrift laat zien dat verschillen tussen scholen het grootst zijn voor burgerschapskennis van leerlingen, en deze verschillen tussen scholen zijn in Nederland groter dan in veel andere landen (Schulz et al., 2017).

Resultaten kunnen ook gebruikt worden om de burgerschapscompetenties van leerlingen te evalueren op het niveau van onderwijsstelsels (Schulz et al., 2017). De resultaten kunnen gebruikt worden om nieuwe beleidsinstrumenten te introduceren, of de effectiviteit van staand beleid te evalueren. De resultaten kunnen ook gebruikt worden om schooleffectiviteit of kenmerken van effectieve scholen te bestuderen (Dijkstra & De la Motte, 2014). De eerste deelstudie van dit proefschrift laat zien dat de resultaten van leerlingen het beleid en discours rond burgerschapsonderwijs reflecteren.

Beoordelingsinstrumenten worden doorgaans ontwikkeld voor een specifiek doel. Gezien de beperkte hoeveelheid instrumenten die beschikbaar zijn om burgerschapscompetenties te beoordelen worden instrumenten echter vaak voor meerdere doelen ingezet. Dit kan leiden tot een discussie over de geschiktheid van

instrumenten voor specifieke doeleinden (Campbell, 1979). Op basis van ICCS 2009 en 2016 concludeerde het Ministerie van Onderwijs, Cultuur en Wetenschap dat de doelstelling voor burgerschapsonderwijs bereikt was, maar dat het burgerschapsonderwijs beter kan (Ministerie van Onderwijs, Cultuur & Wetenschap, 2017). Echter, niet veel later noemde de verantwoordelijke minister het onderzoek "onbruikbaar" (Tweede Kamer, 2018). We zouden echter niet willen stellen dat beoordelingsinstrumenten niet meerdere functies kunnen bedienen. Een docent die bijvoorbeeld toetst of individuele leerlingen een uitleg over de verzorgingsstaat begrepen hebben, zou diezelfde resultaten kunnen gebruiken om te evalueren of de uitleg voldoende helder was. De oplossing is dus eerder een afweging van de geschiktheid van beoordelingsinstrument voor specifieke doelen. De vierde deelstudie in dit proefschrift evalueert de geschiktheid van toetsen en vragenlijsten, portfolio's, *game-based assessment* en vignetten om burgerschapscompetenties te beoordelen, en laat zien dat verschillende typen instrumenten meer of minder geschikt zijn voor bepaalde doeleinden. Het verder uitbreiden van de hoeveelheid beschikbare instrumenten lijkt een zinvolle route (Kerr et al., 2009).

Nieuwe aanpak om burgerschapscompetenties te beoordelen

De tweede en derde deelstudie van dit proefschrift behandelen een bijdrage die dit proefschrift heeft beoogd te realiseren: een nieuw type instrument om burgerschapscompetenties te beoordelen. De ontwikkeling van de rubrics om burgerschap te beoordelen vond plaats gedurende twee jaar, waarin de rubrics met leerlingen, docenten, docentopleiders, studenten, en onderzoekers besproken werden. Bestaande instrumenten voor het beoordelen van burgerschapscompetenties laten veel overeenkomsten zien (vgl. Keating et al., 2010; Schulz et al., 2010; Ten Dam et al., 2011). Deze instrumenten richten zich met name op summatieve beoordeling. Een nieuwe type instrument zou daarom meer gericht kunnen zijn op formatieve beoordeling. De genoemde studies maken bovendien gebruik van instrumenten die ontworpen zijn voor wetenschappelijk onderzoek. Instrumenten die docenten en leerlingen van feedback voorzien zouden kunnen bijdragen aan de toekomstige ontwikkeling van burgerschapscompetenties van leerlingen (Black & William, 1998).

      De tweede deelstudie van dit proefschrift beschrijft twee toepassingen van de rubrics die we ontwikkelden. De eerste manier waarop rubrics gebruikt werden was door leerlingen de rubrics voor te leggen, en hen te vragen welk niveau de leerling het best omschrijft. De resultaten laten zien dat bijna alle leerlingen een niveau konden kiezen, en de meeste leerlingen gaven daar een relevante toelichting

bij. Echter, het percentage leerlingen dat ook een toereikende toelichting gaf voor het niveau dat ze selecteerden wat aanzienlijk lager, met name voor kennis en vaardigheid waarbij één op de vier leerlingen een toereikende toelichting gaf. De tweede manier waarop de rubrics gebruikt werden was door leerlingen een opdracht te laten waarbij ze een reeks reflectievragen beantwoorden. Op basis van hun antwoorden werden de burgerschapskennis en -houding van leerlingen beoordeeld. Inter-beoordelaarsbetrouwbaarheid voor het niveau waarop leerlingen beoordeeld werden was redelijk tot goed.

Omdat de tweede deelstudie liet zien dat studenten vaker een toereikende toelichting gaven bij hun houding, richtte de derde deelstudie zich op zelfbeoordeling van burgerschapshoudingen met behulp van rubrics. De resultaten laten zien dat leerlingen impliciet democratische waarden en principes onderschrijven, en meer expliciet het respecteren van verschillen tussen mensen. Hoewel het hier om een eerste poging gaat, zou deze toepassing in de toekomst docenten, leerlingen en onderzoekers meer inzicht kunnen geven in de burgerschapshoudingen van leerlingen; wat ook docenten en leerlingen kan helpen bij het verder ontwikkelen hiervan. Rubrics lijken dus geschikt voor het beoordelen van burgerschapscompetenties, en voor burgerschapshoudingen in het bijzonder.

Niveaus van burgerschapscompetenties omschrijven

Wanneer burgerschapscompetenties beoordeeld worden is een hogere score over het algemeen 'beter'. Dit blijkt ook uit de eerste deelstudie, waar hogere gemiddelde scores van leerlingen vaak beschouwd worden als minder problematisch in discussies rond burgerschapsonderwijsbeleid. Deze redenering klinkt plausibel wanneer toegepast op burgerschapskennis, waar een dieper begrip van democratische principes en de rol van maatschappelijke instituties zal bijdragen aan beter begrip van democratische processen. Maar is bijvoorbeeld meer vertrouwen hebben in maatschappelijke instituties altijd positief? En wat betekent het wanneer leerlingen aangeven dat ze heel goed zijn (in tegenstelling tot helemaal niet goed) in rekening houden met de wensen van anderen als ze samen een beslissing moeten nemen? Voor deze en andere aspecten van burgerschapscompetenties worden hoger scores van leerlingen over het algemeen ook als 'beter' beschouwd. Door verschillende niveaus van burgerschapscompetenties in rubrics onder woorden te brengen draagt dit instrument bij aan een meer inhoudelijke overweging van wat burgerschapscompetenties inhouden, en hoe deze zich zouden kunnen ontwikkelen.

In eerdere onderzoeken is al getracht verschillende niveaus van burgerschapskennis van elkaar te onderscheiden. Zo omschrijven Schulz en

collega's op basis van de burgerschapskennistoetsen uit ICCS 2009 en 2016 verschillende niveaus van burgerschapskennis (Schulz et al., 2010, 2013, 2017). Wagenaar et al. (2011) vroegen in het kader van de periodieke peiling van het onderwijsniveau burgerschapsonderwijs een groep experts om items voor het meten van burgerschapskennis in te delen naar minimum, voldoende en gevorderd niveau. Daarnaast hebben we gekeken met welke items uit het Burgerschap Meten instrument leerlingen meer of minder moeite hebben (Geboers et al., 2015; Geijsel et al., 2012; Ten Dam et al., 2011). Ten slotte gebruikten we de taxonomie van Bloom om kennisniveaus conceptueel van elkaar te onderscheiden (Bloom, 1956).

Op vergelijkbare wijze hebben we verschillende niveaus van burgerschapshoudingen en -vaardigheden uitgewerkt aan de hand van wat een complexer niveau van beheersing zou kunnen behelzen, zoals een sensitiviteit ontwikkelen voor verschillen tussen mensen in situaties waar dat tot spanningen kan leiden, of ook in een verhitte discussie een consensus kunnen bereiken. Ook hierbij hebben we gekeken op welke items uit het Burgerschap Meten instrument leerlingen hoger of lager scoorden (Geboers et al., 2015; Geijsel et al., 2012; Ten Dam et al., 2011). De inhoudelijke beschrijvingen van de verschillende niveaus worden uitgewerkt in deelstudie twee en drie van dit proefschrift, en de volledige rubrics zijn opgenomen in Appendix B. Waarschijnlijk weerspiegelen de niveaus die hierin worden beschreven ook beter hoe burgerschapscompetenties in het dagelijks leven van leerlingen tot uiting komen, omdat ze ook recht doen aan de complexiteit van de sociale taken waarmee leerlingen in aanraking komen alsook de competenties die nodig zijn om daar effectief mee om te gaan (vgl. Ten Dam & Volman, 2004).

Door hogere niveaus burgerschapscompetenties uit te werken in termen van toenemende complexiteit biedt dit ook meer aanknopingspunten om de ontwikkeling van deze competenties bij leerlingen recht te doen. Eerdere studies laten zien dat de ontwikkelingen van burgerschapshoudingen of -vaardigheden zoals gemeten door vragenlijsten niet lineair verlopen (Geboers et al., 2015; Keating et al., 2010). Door onder woorden te brengen hoe een hoger beheersingsniveau zich zou laten omschrijven, biedt dit leerlingen en docenten ook aanknopingspunten om het leerproces vorm te geven.

Beperkingen

De studies die hier gepresenteerd worden zijn uiteraard ook niet zonder beperkingen. We bespreken er twee hier in meer detail.

Eén meetmoment

De vier niveaus die in de rubrics beschreven worden suggereren verschillende stadia van ontwikkeling die leerlingen doorlopen, op weg naar niveau D. Hoewel de niveaus beschreven zijn met de intentie om het leerproces te ondersteunen, zijn ze niet bedoeld als mijlpalen waarlangs de ontwikkeling van burgerschapscompetenties dient te lopen. De tweede en derde deelstudie gaan dieper in op de implicaties van het gebruik van rubrics om burgerschapscompetenties te beoordelen. De data in deze studies bestaat uit één meetmoment. Het gebruik van rubrics om daadwerkelijk de ontwikkeling van burgerschapscompetenties te volgen en faciliteren verdient verder (longitudinaal) onderzoek. De roep voor longitudinaal onderzoek is niet nieuw, en dergelijke studies zijn schaars (zie bijv. Amnå, 2012; Geboers et al., 2015; Keating et al., 2010; Nieuwelink, 2016).

De rol van taalvaardigheid

Eerdere studies hebben aangetoond dat er een relatie bestaat tussen de ontwikkelingen van taalvaardigheid en burgerschapscompetenties (Eidhof et al., 2017). Gezien het sociale karakter van burgerschap is het niet verrassend dat deze beide begrippen aan elkaar gerelateerd zijn (Starkey, 2005). Burgerschap gaat in essentie om het onderdeel uitmaken van een gemeenschap, en het uitvoeren van sociale taken (hier democratisch handelen, maatschappelijk verantwoord handelen en omgaan met verschillen) vraagt tevens om de taalvaardigheid om dat te kunnen doen. Taalvaardigheid speelt tevens een rol in het beoordelen van burgerschapscompetenties; met name leesvaardigheid en woordenschat kunnen geacht worden een rol te spelen bij het begrijpen van de items waar leerlingen antwoord op geven in een toets of vragenlijst. Deze kwesties zijn niet uniek voor het beoordelen van burgerschapscompetenties, en kunnen ook geacht worden een rol te spelen bij andere beoordelingsinstrumenten die een beroep doen op de lees- of schrijfvaardigheid van leerlingen.

Leesvaardigheid bleek ook een rol te spelen bij het begrip van leerlingen van de rubrics die we ontwikkelden. Bij de pilot van het instrument gaven met name docenten aan dat leerlingen waarschijnlijk moeite zouden hebben om de inhoud van de rubrics te lezen en begrijpen. Dit was tevens een van de redenen waarom we de rubrics tijdens de ontwikkeling met diverse betrokkenen besproken werden, en waarom we in deelstudie twee en drie kozen om het instrument bij oudere leerlingen af te nemen dan in verschillende andere studies naar burgerschapscompetenties. Omdat ruim 90 procent van de leerlingen een relevante toelichting gaf bij de sociale taak en component (t.w. kennis, houding, vaardigheid)

komen we tot de conclusie dat de leerlingen in ieder geval de essentie van de rubrics begrepen. De rol die taalvaardigheid speelt in zowel de werving van burgerschapscompetenties als de beoordeling ervan verdient verdere aandacht, maar lijkt in het geval van rubrics geen grotere rol te spelen dan bij andere beoordelingsinstrumenten.

Zowel de aanpak waarbij leerlingen hun eigen burgerschapscompetenties beoordelen en toelichten, alsook wanneer ze een reflectieopdracht maken die aan de hand van de rubrics beoordeeld worden, werd beïnvloed door de kwaliteit van geschreven antwoorden van leerlingen, en in de meeste gevallen hadden we meer uitgebreide toelichtingen willen zien. Schrijfvaardigheid kan de beoordeling van burgerschapscompetenties verstoren wanneer leerlingen onvoldoende in staat zijn hun antwoorden uit te werken. Wanneer leerlingen gevraagd werden toe te lichten waarom ze hun competenties op een bepaald geplaatst hadden, kopieerden ze vaak delen van de rubrics in plaats van eigen invullingen aan de sociale taak te beschrijven. De antwoorden van leerlingen op de reflectieopdracht varieerden in lengte, en een van de uitdagingen blijft om leerlingen te verleiden hun antwoorden uit te breiden. De belasting die de opdracht doet op het cognitieve vermogen van leerlingen zou hier ook een rol gespeeld kunnen hebben. Leerlingen hadden er echter geen eigen belang bij de opdracht zo goed mogelijk te maken, wat deze problemen in onze studies mogelijke heeft vergroot.

Deze beperkingen zouden geleidt kunnen hebben tot een onderschatting van de burgerschapscompetenties van leerlingen. Voor de toepassing van rubrics om burgerschapscompetenties te beoordelen betekent dit dat leerlingen vooral ondersteuning verdienen bij het toelichten van hun beoordeling of antwoorden. Vervolgonderzoek zou tevens kunnen kijken of aanpakken die minder afhankelijk zijn van geschreven antwoorden van leerlingen mogelijk zijn.

Implicaties

Neem de beleidscontext rond burgerschapsonderwijs in ogenschouw

Burgerschapsonderwijs staat de afgelopen decennia in veel landen op de beleidsagenda. De eerste deelstudie van dit proefschrift laat zien dat het beleidsprobleem vooral in de nationale context beschouwd wordt, en hierbij komen zowel overeenkomsten als verschillen tussen landen naar voren. Het is van belang oog te hebben voor deze beleidsontwikkelingen, omdat ze niet alleen van invloed zijn op toekomstig beleid, maar ook iets zeggen over de burgerschapscompetenties van leerlingen (Hahn, 1999). Westheimer en Kahne (2004) laten zien dat verschillende doelen nagestreefd kunnen worden in het kader van

burgerschapsonderwijs. Leenders, Veugelers en de Kat (2008) laten zien dat deze verschillende doelen ook onder docenten terug te vinden zijn. Het is van belang deze discussies mee te nemen in het beoordelen van burgerschapscompetenties om de normatieve validiteit van beoordelingen te realiseren (Biesta, 2010). Zoals in de vierde deelstudie uitgewerkt, verschillen typen beoordelingsinstrument in de mate waarin zij het waarderen van verschillende concepties van burgerschap mogelijk maken. Bij het kiezen van een beoordelingsinstrument is het daarom ook van belang om rekening te houden met de conceptuele uitwerking van burgerschapscompetenties.

Een geschikte beoordelingsaanpak kiezen
Het beoordelen van burgerschapscompetenties is geen eenvoudige kwestie gebleken. De studies die als onderdeel van dit proefschrift zijn uitgevoerd hebben verschillende implicaties voor verdere ontwikkeling van het beoordelen van burgerschapscompetenties. Op basis van de deelstudies komen we tot een aantal aanbevelingen. Hierbij kan onderscheid gemaakt worden tussen het beoordelen van burgerschapskennis, -houdingen, en -vaardigheden.

Beoordelen van burgerschapskennis
Burgerschapskennis van leerlingen kan op verschillende manieren beoordeeld worden, en het gebruik van rubrics bleek redelijk tot goed betrouwbaar. Wanneer rubrics gebruikt werden voor zelfbeoordeling gaf maar één op de vier leerlingen een toereikende toelichting bij het zelfgekozen niveau, waardoor inzicht in hun daadwerkelijk kennisniveau beperkt is. Toetsen lijken het meest geschikte instrument te zijn om burgerschapskennis te beoordelen. Hierbij bedoelen we echter niet te zeggen dat het hierbij alleen om multiple-choice toetsen gaat. Hoewel deze zeker hun voordelen hebben, kunnen kennisvragen op allerlei manieren geformuleerd worden, variërend van multiple-choice tot het schrijven van een opstel. Meer open-geformuleerde vragen zouden daarbij ook met rubrics beoordeeld kunnen worden, mits leerlingen hun antwoorden voldoende uitwerken.

Beoordelen van burgerschapshoudingen
Burgerschapshoudingen worden in de regel gemeten met behulp van een reeks items waarbij leerlingen antwoorden op een Likertschaal. Deze items zijn gemakkelijk af te nemen, en genereren snel een gestandaardiseerde uitkomst. De tweede en met name de derde deelstudie in dit proefschrift laten zien dat burgerschapshoudingen ook beoordeeld kunnen worden met behulp van rubrics, en dat een dergelijke benadering een meer holistische beoordeling van

burgerschapshoudingen mogelijk maakt waarbij meer recht wordt gedaan aan de manier waarop leerlingen burgerschapshoudingen ontwikkelen. Er is geen duidelijk 'beste instrument' om burgerschapshoudingen te beoordelen. Vragenlijsten bieden diverse schalen die bewezen hebben burgerschapshoudingen van leerlingen in kaart te kunnen brengen. Binnen de gesloten benadering waarbinnen burgerschapshoudingen uitgewerkt worden bieden vragenlijsten een effectieve beoordeling van burgerschapshoudingen. Wanneer het doel is een rijker beeld te schetsen van de burgerschapshoudingen van leerlingen, lijken rubrics een zinvolle aanvulling op het bestaand instrumentarium.

Beoordeling van burgerschapsvaardigheden
Bij het gebruik van rubrics gaf slecht één op de vier leerlingen een toereikende toelichting bij het niveau van burgerschapsvaardigheid dat zij zichzelf toekenden. Dit beperkt de mogelijkheden om uitspraken te doen over het daadwerkelijke niveau van burgerschapsvaardigheid dat zij beheersen. In enquêtes waarbij vragenlijsten gebruikt worden om burgerschapsvaardigheden te beoordelen wordt veelal gevraagd aan leerlingen in hoeverre zij zichzelf in staat achten om een bepaalde taak uit te voeren ('*self-efficacy*'). Zowel zelfevaluatie aan de hand van rubrics als aan de hand van items op een Likertschaal lijken gevoelig voor andere invloeden zoals het Dunning-Kruger effect of assertiviteit van leerlingen, waardoor de beoordeling minder betrouwbaar wordt. Voor het beoordelen van burgerschapsvaardigheden lijkt verdere ontwikkeling van beoordelingsinstrumenten zinvol, met name waar het gaat om benaderingen waarbij leerlingen authentieke taken uitvoeren op basis waarvan hun vaardigheden beoordeeld kunnen worden (zie deelstudie vier; vgl. Roelofs & Sluijter, 2016).

Conclusie

Dit proefschrift biedt een overzicht van verschillende manieren om burgerschapscompetenties te beoordelen. Burgerschapscompetenties bestrijken een breed palet aan thema's, zoals kennis hebben van democratische besluitvorming of houdingen ten aanzien van sociale verschillen. Beoordelingsinstrumenten richten zich daarom in de meeste gevallen op een reeks van aspecten. Bovendien kunnen burgerschapscompetenties op verschillende manieren beoordeeld worden, en voor verschillende doeleinden. In dit proefschrift wordt onder andere een aantal criteria uitwerkt aan de hand waarvan toetsen en vragenlijsten, portfolio's, *game-based assessment* en vignetten geëvalueerd worden. Een nieuw ontwikkelde set rubrics om burgerschapscompetenties te beoordelen is getest en geëvalueerd. Het

bestaande instrumentarium maakt met name gebruik van toetsen en vragenlijsten, en rubrics lijken hier een zinvolle aanvulling op te bieden door meer inhoudelijk inzicht te genereren in de burgerschapscompetenties van leerlingen. Het lijkt zinvol het beschikbare instrumentarium verder uit te breiden, zodat op basis van de doelen die beoogd worden het meest geschikte instrument gekozen kan worden.

DANKWOORD

"Dames en heren, het gebouw gaat over vijftien minuten sluiten. U wordt vriendelijk verzocht uw werkzaamheden af te ronden, en het gebouw voor tien uur te verlaten. --- Ladies and gentlemen, the building will close in fifteen minutes. You are kindly requested to finish what you are doing, and leave the building before ten o'clock." Zo eindigde menig werkdag aan de UvA. Misschien dat ik fulltime werken soms iets te letterlijk nam, maar als je van je werk je hobby hebt gemaakt is dat absoluut geen straf. Mijn beeld van te voren – voor zover ik dat al had – was dat promoveren vooral een solitaire aangelegenheid zou zijn, maar het is de afgelopen jaren een heel sociale bedoening gebleken. Daarin heb ik met vele mooie mensen mogen samenwerken.

In de eerste plaats ben ik mijn promotoren ontzettend dankbaar: Anne Bert Dijkstra, Sjoerd Karsten, en – tijdens de eerste twee jaar – Geert ten Dam. Anne Bert, ik weet nog goed dat ik dacht dat ik voor een tweede ronde van het sollicitatiegesprek met je afsprak op Schiphol, omdat het op die manier goed tussen je andere afspraken paste. Al snel bleek dat het niet om een sollicitatiegesprek ging, maar dat ik blijkbaar de positie al aangeboden kreeg, en dat we aan het afstemmen waren hoe dat er ongeveer uit zou komen te zien. Ik heb vanaf het begin bewondering gehad voor je werklust en kennis van zaken. We zitten vaak op dezelfde golflengte, en ik kijk terug en vooruit op een geweldige samenwerking. Dank ook voor alle deuren die je daarbij voor me opende.

Sjoerd, dank dat je het stokje van Geert wilde overnemen toen zij collegevoorzitter werd. Ik schreef mijn masterscriptie al onder jouw begeleiding, en vrienden snappen nog steeds niet waar ik het over had wanneer ik zei dat het leuk was om aan mijn scriptie te werken. Ik had er dus alle vertrouwen in dat het schrijven van mijn proefschrift onder jouw begeleiding ook wel goed zou komen. De dynamiek die ontstond wanneer we met zijn drieën afspraken zorgde dat ik altijd met een glimlach onze afspraken in- en uitging. Hopelijk heb je nog veel printverzoeken de komende jaren, want het is nooit saai als je langskomt.

Geert, je moest helaas stoppen als mijn tweede promotor toen je de gevraagd werd als voorzitter van het college van bestuur van de universiteit. Gelukkig bleven we wel met elkaar samenwerken, onder andere bij het vak Onderwijsbeleid wat we samen met Anne Bert in de master geven. Het is indrukkend – maar in het geheel niet verrassend – om te zien hoe goed je ook de rol van collegevoorzitter weer schijnbaar moeiteloos vervult. Ik ben blijvend onder de indruk van je kwaliteiten, en dankbaar voor het voorbeeld dat je me geeft.

Behalve mijn promotoren heb ik vele mensen mogen vragen om mee te denken met de rubrics die ik voor dit proefschrift ontwikkeld heb. Ik ben de vele docenten die in de verschillende fases van ontwikkeling en afname van de rubrics geholpen hebben dankbaar voor de tijd die zij vrijgemaakt hebben: Aniek, Anke, Anna, Betty, Chiel, Chris, David, Dick, Eddie, Emile, Emily, Erik, Gill, Ieke, Ilse, Ilse, Ivo, Jonneke, Josephine, Henri, Jeroen, Koen, Krista, Laurie, Lida, Linda, Lotte, Malty, Mano, Marian, Marlies, Paul, Paul, Pieter, Puck, Renske, Sanneke, Tamara, Thijs, Thomas, en Zohair ontzettend bedankt. Het was een genoegen om met zoveel enthousiaste, geïnteresseerde docenten samen te werken, en daarbij ook veel te zien van de onderwijspraktijk waar het uiteindelijk allemaal om te doen is. Daniëlle, Kevin en Maaike bedankt voor jullie hulp bij het beoordelen van het werk van de leerlingen en ook jullie feedback bij het werken met de rubrics.

Ik was de afgelopen jaren veel op de universiteit, en werk daar nog steeds met veel plezier. Dat komt ook door het team van collega's waar ik deel van uitmaak. Monique, je gaf de afgelopen jaren leiding aan onze onderzoeksgroep, en inmiddels aan de afdeling. Ik was als student al overtuigd van je kennis en kunde, en dat werd in mijn ervaring als collega alleen maar bevestigd. Dat we zo'n betrokken onderzoeksgroep hebben is wat mij betreft ook jouw verdienste.

Frank en Willemijn, we begonnen samen op 1 september 2013 in de kelder van de UvA aan onze avonturen. Drie heel verschillende mensen op drie heel vergelijkbare posities. Ik heb genoten van onze samenwerking en hoop dat we elkaar zullen blijven tegenkomen.

Na ons begin aan de UvA kwamen Bram en Hessel al snel om de hoek kijken als de promovendi die ons voorgingen. Toen we na de afdelingsverhuizing ook bij elkaar op de kamer kwamen werd het al snel gezellig in 'de burgerschapskamer'. Van jullie weet ik zeker dat we elkaar nog veelvuldig gaan treffen; al zal het waarschijnlijk niet meer zijn om te tafeltennissen.

Al is wetenschappelijk onderzoek vaak een langzaam proces, als je maar lang genoeg honkvast blijft wordt je vanzelf de langstzittende collega op de kamer. Ik ben inmiddels de nestor van D8.06. Met Anne, Ati, Daury, Ellen, Hester, Jitske, Lianne, Lina, Malin, Manja, Mayke, Monique, Rutmer, Semiha, en Suzan was en blijft het een heerlijke werkplek.

Anke en David, mijn paranimfen. Allereerst bedankt dat jullie het op je wilde nemen om nog iets meer een feestje van mijn promotie te maken. Anke, we begonnen in ongeveer dezelfde periode met ons werk aan de UvA, en vanaf het begin was duidelijk dat je een ontzettend enthousiaste en fijne collega bent. Mooi dat je de klap die je ruim een jaar geleden kreeg inmiddels weer te boven bent, en we gaan nog vele mooie dingen bereiken en veel lekkere koffie drinken. David, je

hebt volgens mij nooit begrepen waar ik het over had wanneer ik je vertelde dat scriptie schrijven en promoveren ontzettend leuk is, maar we begrijpen elkaar heel goed als het gaat om het doen van werk waar je plezier in hebt en waar je je door uitgedaagd voelt. 'Mijn scriptietje' is nu eindelijk af. Ik zou kunnen zeggen dat ik de komende periode weer meer tijd heb om af te spreken, maar het zou zomaar kunnen dat er weer wat nieuw hooi op de vork komt. Maar lukt het vast nog wel regelmatig om een kort potje te spelen.

Als voorzitter van de promovendiraad schreef ik al eens een stukje over dat promoveren nog wel eens als stressvol ervaren wordt, en het is dan fijn als het leven buiten de universiteit stabiliteit biedt. Daaraan heb ik geen gebrek gehad. Papa en mama, ik heb me altijd gesteund gevoeld in de keuzes die ik gemaakt heb en jullie staan altijd voor me klaar. Ik zeg het niet vaak genoeg: ik houd van jullie. Maarten en Marijke, in de tijd dat ik aan mijn proefschrift werkte hebben jullie een gezin gesticht. Mooi om te zien hoe mijn kleine zusje groot is geworden, en hoe jullie samen gelukkig zijn. Iris, jij houdt het inmiddels al bijna een jaar uit met al mijn kwaliteiten en gebreken, en ik ben blij dat ik je in mijn leven heb.

ABOUT THE AUTHOR

Remmert Daas was born on 16 August 1985 in Hoorn, the Netherlands. After secondary school in Enkhuizen, he completed a bachelor's degree in teaching English at the Amsterdam University of Applied Sciences. In the following three years he taught in senior vocational education. He then studied Educational Sciences at the University of Amsterdam, and wrote his master's thesis on citizenship education in the Netherlands, England and Sweden. His PhD research focused on different methods for assessment of citizenship competences, inter alia by developing and evaluating a rubric instrument to assess citizenship competences. He now works as a Postdoc researcher at the University of Amsterdam and is engaged in several research projects regarding evaluation of students' citizenship competences and other skills. He also teaches several courses on education policy.