Contemplating modes of assessing citizenship competences

Daas, R.; ten Dam, G.; Dijkstra, A.B.

Link to publication

Contemplating modes of assessing citizenship competences

Remmert Daas*, Geert ten Dam, Anne Bert Dijkstra

Department of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

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.

© 2016 Elsevier Ltd. All rights reserved.

1. 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

* Corresponding author at: Department of Child Development and Education, University of Amsterdam, Postbus 15776, 1001 NG, Amsterdam, The Netherlands.
  E-mail address: r.j.m.daas@uva.nl (R. Daas).

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 incentivise 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 endeavour. 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?*

2. 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.

2.1. 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.

2.2. 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.

2.3. 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 socio-cultural 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.

2.4. 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.

3. 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.

4. Results

4.1. 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 Cool[5−18] 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 optimise 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 behaviour 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 endeavour. 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).

4.2. 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 programmes that had flexible or individualised 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.

4.3. 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[1] 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[2] 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 behaviour 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 behaviour 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 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 behaviour 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 behaviour 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.

4.4. 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-

---

[1] https://communityplanit.org/.
[2] http://globalconflicts.eu/.

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.

5. 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 endeavour.

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 behaviour – 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 behaviour can be transferred to real life situations. The largest drawback however, is that for game-based assessment of citizenship competences to realise 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.

References

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

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

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. (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.

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.

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.

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.

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

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.

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

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.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27, 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.). ([74_TD$DIFF]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.

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/B0-08-043076-7/02367-6.

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.

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.

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, Belgium: Eurydice.

Fischman, & 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.

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

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

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.

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.

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

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., Vilalba, C., & Saisana, M. (2012). *The 2011 civic competence composite indicaor (CCCI-2)*. Luxembourg, Luxembourg: Publications Office of the European Union.

Huisman, J., Pijls, T., Van Hoeij, J., Van Voorst van Beest, K., Boonaerts, Y., & Lens, M. (2003). *Portfolio sociale competenties: Primair onderwijs, mbo, 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.

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.

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., 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.

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.

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.

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

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

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

Ledoux, G., Meijer, J., Vander 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.

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

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.

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.

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

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

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.

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.

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.

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

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

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.

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

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

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., 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., 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.

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.

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.

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.

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.

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., & 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.

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.

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.