



## UvA-DARE (Digital Academic Repository)

### An open-ended approach to evaluating students' citizenship competences: The use of rubrics

Daas, R.J.M.; Dijkstra, A.B.; Karsten, S.; ten Dam, G.T.M.

**Publication date**

2023

**Document Version**

Final published version

**Published in**

Education, Citizenship and Social Justice

[Link to publication](#)

**Citation for published version (APA):**

Daas, R. J. M., Dijkstra, A. B., Karsten, S., & ten Dam, G. T. M. (2023). An open-ended approach to evaluating students' citizenship competences: The use of rubrics. *Education, Citizenship and Social Justice, Online First*, 1-16. Advance online publication. <http://10.1177/17461979231186028>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# An open-ended approach to evaluating students' citizenship competences: The use of rubrics

Education, Citizenship and  
Social Justice  
1–16

© The Author(s) 2023



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/17461979231186028  
journals.sagepub.com/home/esj



**Remmert Daas** 

University of Amsterdam, The Netherlands

**Anne Bert Dijkstra**

University of Amsterdam, The Netherlands  
National Inspectorate of Education of the Netherlands

**Sjoerd Karsten**

University of Amsterdam, The Netherlands

**Geert ten Dam** 

University of Amsterdam, The Netherlands

## Abstract

In this paper, we develop an open-ended approach to evaluating students' citizenship competences. We aim to give students the opportunity to describe what citizenship means for them in personally relevant contexts. We developed three rubrics relevant to students' citizenship in daily life. Students in grade 10 and 11 ( $M_{\text{age}} = 16$ ) evaluated their competences or completed an assignment which was assessed using rubrics. The results show that for both approaches, the majority of students were able to provide relevant input pertaining to their citizenship competences. However, students' explanations were often brief, limiting the personal context they provided and the extent to which they demonstrated higher levels of proficiency. This study shows that employing rubrics for an open-ended approach to assessing citizenship competences shows promise in allowing students to share and elaborate upon their experiences and viewpoints, but more focus is needed on improving the quantity and quality of student input.

## Keywords

citizenship competences, evaluation, rubrics, self-evaluation, social tasks

---

### Corresponding author:

Remmert Daas, Department of Child Development and Education, University of Amsterdam, Postbus 15780, 1001 NG Amsterdam, The Netherlands.

Email: r.j.m.daas@uva.nl

## Introduction

In order to participate effectively in a democratic society, young people need citizenship competences. This includes knowledge of democratic decision-making, the skills to talk about perspectives on various civic issues and an attitude supportive of democratic rights and responsibilities (Council of Europe, 2016; Ten Dam et al., 2011). Most Western-European countries have implemented legislation for schools to promote development of students' citizenship competences (Eurydice, 2017). This has also been the case in the Netherlands, where in 2005/2006, the Dutch government laid down the promotion of citizenship in educational legislation and in the core objectives of primary and secondary education. The statutory citizenship obligation was formulated in rather unspecified terms, and schools could determine its content, nature, and extent. The legislation on citizenship education in primary and secondary education was recently expanded. Schools should actively promote knowledge, attitudes and skills around the basic values of the democratic rule of law, and the school culture should also be consistent with it (De Groot et al., 2022). This statutory citizenship obligation raises the question how insight into students' development of citizenship competences can be generated.

Up to now, tests and questionnaires are the most frequently used instruments to evaluate students' citizenship competences. Although the use of closed-ended questions (i.e. multiple choice items) has facilitated comparisons between students, school and even countries, it offers little insight into *why* students selected a particular answer. Because of the closed format, students are unable to explain whether they feel strongly about a particular civic issue, and the extent to which they consider issues pertaining to citizenship important to themselves and others. Tests and questionnaires are therefore arguably less suited to reflect the citizenship practices deemed relevant by students, nor do these provide an understanding of the students' reasoning or the arguments students use to explain why they agree or disagree with a particular statement (Daas et al., 2016).

Open-ended questions provide students with the opportunity to explain their reasoning in response to various questions. These have for example been used in interviewing students about their citizenship attitudes (Nieuwelink et al., 2018; Vaessen et al., 2022). However, open-ended questions allow for a wide range of directions for students' responses, and evaluating the level of students' citizenship competences based on their answers has proven a challenge. An approach receiving considerable, and increasing, attention to this end is the use of rubrics (Dawson, 2017; Jonsson and Svingby, 2007; Panadero and Jonsson, 2013; Pancorbo et al., 2021). These allow assessors to evaluate students' responses using distinctly described levels of proficiency.

In this article, we focus on rubrics as a novel approach to the evaluation of citizenship competences. Using open-ended questions and rubrics, we aim to give students the opportunity to describe what citizenship can mean for them in contexts that are relevant to their lives. Asking students to explain their views may allow to uncover students' personal (mis)conceptions in relation to citizenship. From an educational point of view, such understanding supports teachers to foster citizenship competences in ways that are meaningful to their students.

We test two approaches to open-ended questions and rubrics. Our first approach concerns students' self-evaluation of their citizenship competences. Students use a rubric to evaluate their own competences. Our second approach presents students with an assignment for which they answer a series of questions. We use a rubric to evaluate students' work. In this article both the use of rubrics through students' self-evaluation and the assessment of students' written work are considered. Our main research question is: *What are the advantages and disadvantages of using rubrics to evaluate the citizenship competences of students?* Following the two distinguished uses of rubrics, this can be broken down into the following research questions:

- Do rubrics support student self-evaluation of citizenship competences?
- Do rubrics support the evaluation of citizenship competences based on student work?

## Citizenship competences in the context of students' daily lives

For some time now, strengthening students' citizenship competences through education has become an important theme in political and public debate, as well as in scientific research. In many Western-European countries, citizenship education has become a statutory duty for schools (Eurydice, 2017). Increasingly, democratic societies feel a need to reinforce the citizenship competences of their members. This need arises from a growing recognition that democracy is essentially fragile and depends on engagement of citizens in both formal participation (e.g. voting) as well as more generally supporting a democratic culture (Osler and Starkey, 2006). Moreover, there is a desire to increase social cohesion (Oser and Veugelers, 2008), that is, to harden the 'social glue' that keeps society together and ensures that citizens feel committed to each other and are involved in society.

Engagement of citizens in various forms of social life takes place both within the domain of politics and within the larger social domain. It involves knowledge of, for example, civic institutions and the separation of powers, skills to take an active part in political and social life and positive attitudes towards others and society as a whole (Schulz et al., 2016). In other words, citizenship competences refer to the ability to live together in different roles, in relation to other individuals and groups and as citizens interacting with government (Rychen and Salganik, 2003). These aims can be regarded as globally relevant, evidenced by the formulation of sustainable development goal 4.7 by the United Nations (2015).

However, despite this seeming consensus on what citizenship competences refer to, different views on what constitutes 'good citizenship' coexist (Eidhof et al., 2016; Westheimer and Kahne, 2004), and these reflect both normative understanding of a just society and changes in the context in which citizenship is considered (Abowitz and Harnish, 2006). Several overviews have shown how citizenship competences refer to a wide range of civic domains (Cogan and Morris, 2001; Council of Europe, 2016; Schulz et al., 2016). This means that assessment of citizenship competences requires further demarcation of which behaviours in social domains or situations are deemed relevant to the assessment. The competences pertaining to these situations are considered distinct competences with regards to specific kinds of behaviours, because there is little evidence that citizenship competences can be considered one composite measure applicable to all domains (cf. Hoskins et al., 2012). In other words, the plurality of citizenship competences implies we are referring to multiple qualities, rather than a single one.

When considered citizenship competences of young people, another relevant question is whether these competences refer to citizenship behaviours when students have become adults, or if students are already considered citizens in their own right. Lawy and Biesta (2006) advocate citizenship should extend to children and young people and for citizenship education to be effective, it should recognize meaningful practices of citizenship in which young people engage. This viewpoint thereby aims at a more personally meaningful as well as educationally and socially relevant conception of citizenship of young people.

In line with this perspective, this study builds on a conceptualization of Ten Dam and Volman, 2007; Ten Dam et al., 2011) that focuses on citizenship competences as relevant to the *daily lives of students*. To evaluate the capabilities of students to take an active part in a democratic society, young people's citizenship is considered in four exemplary social tasks: acting democratically, acting socially responsible, dealing with conflicts and dealing with differences. Adequate performance of these tasks depends on students' relevant knowledge, attitudes and skills. Citizenship

competences can thus be measured by assessing students' knowledge, attitudes and skills pertaining to each social task.

The social task acting democratically refers to acceptance of and contribution to a democratic society, for example, knowing and acting in accordance with democratic principles. Acting socially responsible refers to taking shared responsibility for the communities to which one belongs, for example, wanting to uphold social justice. Dealing with conflicts refers to handling of situations of conflict or conflict of interest to which the adolescent is a party, for example, seeking win-win solutions. Dealing with differences refers to handling of social, cultural, religious, and outward differences, for example, having a desire to learn other people's opinions and lifestyles. These tasks and their underlying knowledge, attitudes and skills are further described in Ten Dam and Volman (2007) and Ten Dam et al. (2011).

The emphasis on social tasks underlines the relevance of the context in which citizenship competences are employed. Because our study is conducted in the Netherlands, this concerns the societal contexts relevant to Dutch students. However, the themes that play a role here – such as diversity, democracy, equality and inequality and social trust – are fundamental to modern democratic society, as can be seen from the attention to these issues in other Western-European, democratic societies (Council of Europe, 2016; Eurydice, 2017). Although students in secondary education will generally not have reached voting age, they are regarded as young citizens, that is, members of a community or society, since they interact with their socio-political environment and with other individuals and groups around them.

Promoting citizenship competences is a statutory responsibility for Dutch schools, the objectives of which were first stipulated in legislation in 2005 and expanded in 2021 (De Groot et al., 2022). Citizenship education is a schoolwide responsibility. However, national learning aims to citizenship show particularly close ties to the subjects of civics, history and geography (Nieuwelink and Oostdam, 2021). Civics has been a school subject in upper secondary education since 1968 and includes the study of topics such as 'parliamentary democracy' and 'the diverse society'. Civics examinations are the responsibility of individual schools. With respect to citizenship education, schools are free to choose if and how to assess whether students have met the goals stipulated in the national curriculum. Recent changes in legislation require schools to gather insight into students' citizenship competences. This is one of the areas in which schools were found lacking in earlier national evaluations (Inspectorate of Education, 2016).

Most instruments available to researchers and schools use tests and questionnaires with closed-ended questions. It is worthwhile to consider alternative approaches, with more emphasis on providing students with opportunities to elaborate their answers in relation to their considerations and the contexts in which their citizenship competences develop. Rubrics can be a suitable tool for the assessment of 'complex performances' (Jonsson and Svingby, 2007) such as citizenship competences.

## **Rubrics**

Rubrics aim to provide a scoring tool of authentic or complex student work (Jonsson and Svingby, 2007). Essentially, a rubric is a matrix with two axes: the assessment criteria, and the levels of proficiency. Each of the cells in this matrix specifies a description of the expected performance. The level of detail in the performance descriptions can range from broad descriptions within 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'). The number of levels of proficiency is typically three to five (Brookhart, 2018).

Holistic rubrics typically consist of only one or few criteria, whereas analytic rubric can vary greatly in the number of scoring criteria.

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). Although rubrics are assumed to promote learning by making expectations and criteria explicit and by facilitating feedback and self-assessment, reviews by Jonsson and Svingby (2007) and Panadero and Jonsson (2013) find little evidence available that rubrics have these benefits on student learning. A more recent review by Brookhart and Chen (2015) concludes the evidence is ‘promising but not sufficient for establishing that using rubrics cause increased performance’ (p. 363).

The information provided by rubrics and the room they offer to assess a variety of student behaviours, means evaluation is inherently embedded in context – especially when rubrics are directed at a specific task (Dawson, 2017). Students can be directed to reflect on their acting in these contexts, by asking them to consider how they perceived the context. These qualities appear to make rubrics particularly well suited for meaningful evaluation of citizenship competences, because they provide students with opportunities to elaborate upon their personal feelings and beliefs concerning citizenship – characteristics that closed-ended questions are generally less suited for (Daas et al., 2016).

These benefits imply rubrics can be used in (at least) two ways: students may use rubrics for self-evaluation, or rubrics may be used to assess students’ work. To the best of our knowledge, no research has yet been conducted into the use of rubrics to evaluate citizenship competences. In this article we investigate both applications of rubrics (i.e. self-evaluation and evaluation by assessors) on citizenship knowledge, attitudes and skills.

## Method

### *Development of rubrics for citizenship competences*

For the purpose of this study, we designed a set of rubrics aimed at evaluating student citizenship competences. Based on the conceptual framework developed by Ten Dam and Volman (2007) described above, we developed three rubrics for three exemplary social tasks relevant to student citizenship in daily life: acting democratically, acting socially responsible and dealing with differences. Although we initially also developed a rubric for the social task dealing with conflicts, the teachers with whom we collaborated indicated that dealing with conflicts is not an explicit part of their curriculum. Therefore, we decided not to develop this rubric any further.

Each rubric describes the social task in terms of knowledge, attitude and skills (i.e. the assessment criteria of the rubric) at four levels: A through D. For each cell in each of the rubrics, we developed a general description and several examples of typical knowledge, attitudes or skills. The descriptions were based on conceptualizations of levels of citizenship competences following earlier measurements of student citizenship competences (cf. Geboers et al., 2015; Keating et al., 2010; Schulz et al., 2008; Ten Dam et al., 2011; Torney-Purta et al., 2001). The rubrics were developed over the course of 2 years, during which researchers, teachers and students were consulted for input and feedback.

The draft versions of the rubrics were discussed with eight Civics teachers working in Dutch secondary schools and with six panels of students in the second, third and fourth years of secondary education (grades 8 through 10; mean age 13.5 through 15.5 years old). The discussions with both the teachers and the students showed that younger students (aged 13 or 14 years) had considerable difficulty understanding the rubrics. They appeared to consider the examples included in the rubrics as a checklist and had difficulty relating the contents of the rubrics to their own experiences as

|                            | Level A | Level B  | Level C | Level D |
|----------------------------|---------|--|---------|---------|
| Knowledge                  | ...     | ...  | ...     | ...     |
| Attitude<br>(I believe...) | ...     | <p>You assume everyone should abide by the same rules. You believe that while people differ, everyone is equal.</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> <li>• <i>I believe that men and women deserve equal opportunities.</i></li> <li>• <i>I believe that people should be able to be who they want to be, as long as they abide by the rules.</i></li> <li>• <i>I believe that it is normal to respect other religions, genders and sexual orientations.</i></li> </ul> | ...     | ...     |
| Skills                     | ...     | ...  | ...     | ...     |

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

young citizens. This would make it difficult for rubrics to enable students to provide explanations by using examples from their own lives, which is what we were after. Because our study aimed at investigating the viability of using rubrics to assess citizenship competences, we decided to focus on students older than 14 years and made several layout changes to discourage students from treating the rubrics as a checklist. An added benefit of focussing on students of 15 years and over is that in the Dutch education system students in the fourth year of general secondary education (grade 10) have ‘Civics’ classes and students in the first year of vocational tertiary education (grade 11) have ‘Career and Citizenship’ classes, which meant that we could cooperate with the teachers who taught these classes for the remainder of the study. These findings resulted in a revised version of the rubrics.

We tested the revised rubrics in a pilot involving nine Civics and Career and Citizenship teachers in grade 10 of a general secondary school or grade 11 of a tertiary vocational school. Students from 11 classes evaluated their citizenship competences using one of the rubrics. Students from nine other classes completed an assignment in which they were asked to answer a set of open-ended questions; their answers were assessed using the knowledge and attitude dimensions from one of the rubrics. This pilot provided valuable input, causing us to change some of the phrasings to clarify the structure and differences between levels, which is considered to be one of the most challenging aspects of designing rubrics (Pancorbo et al., 2021; Reddy and Andrade, 2010; Tierney and Simon, 2004). This testing provided input to re-evaluate whether the examples in the rubrics matched the corresponding level of competence. Because the pilot results were used to revise the rubrics, the data from these students were not included in the main study.

These steps resulted in three rubrics, covering acting democratically, acting socially responsible and dealing with differences, respectively. Each rubric consists of three dimensions (knowledge, attitudes and skills) and four levels (A, B, C, D). Each cell in the rubrics contains a general description and three examples of the corresponding knowledge, attitudes and skills. Figure 1 shows the cell representing attitudes towards dealing with differences at level B as an example. The complete rubrics can be found in Appendix A.

### Participants

To evaluate the final versions of the rubrics we approached teachers of Civics and Career and Citizenship to participate in our study. These teachers were contacted through the authors’ networks, as well as by placing adverts on websites and online discussion groups on (citizenship) education. Thirteen Civics teachers of grade 10 students in general secondary education and six

**Table 1.** Participant data description.

| Level of education | N   | Mean age | % female | Self-evaluation | Reflection assignment |
|--------------------|-----|----------|----------|-----------------|-----------------------|
| Secondary          | 456 | 15.4     | 50%      | 383             | 73                    |
| Vocational         | 260 | 17.4     | 57%      | 218             | 42                    |
| Total              | 716 | 16.2     | 53%      | 601             | 115                   |

Career and Citizenship teachers of grade 11 students in vocational tertiary education participated in the main study. These teachers incorporated either the self-evaluation or the reflection assignment (see below) into one or more of their lessons at some time between September 2016 and February 2017. The teachers selected one of the three rubrics which matched the topic that they were teaching at that time. For example, if ‘the diverse society’ (one of the common themes in Civics) was the topic, the teacher would select the rubric on dealing with differences. All students had to complete either the self-evaluation ( $N=601$ ) or the reflection assignment ( $N=115$ ) as part of the regular lessons. Students were informed that the assignment was 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. Because teachers selected the social task most relevant to their curriculum, respondents are not distributed equally over the different social tasks. As part of the self-evaluation task, 135 students completed the assignment on acting democratically, 172 students on acting socially responsibility and 294 students on dealing with differences. As part of the reflection assignment task, 42 students completed the assignment on acting socially responsibility and 73 students on dealing with differences. In total, 716 students took part in the study (see Table 1).

## Materials

### *Self-evaluation task*

For the self-evaluation task, students evaluated their knowledge, attitudes and skills pertaining to acting democratically, acting socially responsible, or dealing with differences. The assignment had to be completed in class and students generally finished it within 30 minutes. Students were presented with the rubric 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 and ‘For example. . .’ for each.

The explanations that students provided for their answers 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 for the selected level? The first question is intended to assess whether students understood the assignment and the topic. The second question is intended to assess whether students were able to accurately evaluate their own citizenship competences using the rubric and provide sufficient justification.

In order to assess inter-rater reliability of the rubrics, three masters students of educational sciences were instructed on the structure, contents and aims of the rubrics by the first author in an interactive meeting of 45 minutes. The three students and first authors rated students’ explanations. The explanations by the first 40 students were assessed by all four raters, after which differences



in grading were discussed to reach consensus. The other explanations were assessed by three raters: the first author and two of the other raters in rotation. The final assessment of the explanation's relevance and adequacy was based on rater majority. Inter-rater agreement was calculated to determine the reliability of scoring. Interrater reliability of rubrics is commonly expressed as the percentage of agreement between raters (Jonsson and Svingby, 2007). Because inter-rater agreement does not correct for chance, we calculated Cohen's kappa for each pair of raters. Cohen's kappa adjusts for chance agreement, for instance if raters are overall more likely to answer 'yes' than 'no', inter-rater agreement is likely to increase while Cohen's kappa will correct for this likelihood (Landis and Koch, 1977).

### *Reflection assignment task*

For the reflection assignment, students completed an assignment covering either acting socially responsible or dealing with differences. We developed two assignments which followed a lesson or series of lessons on the topic (the assignments are included in Appendix B). The reflection assignment on acting socially responsible focused on the perception of social inequalities and included a set of questions (e.g. 'What positive consequences can social differences have?') that students should answer. The reflection assignment covering dealing with differences focused on cultural differences and prejudices about immigrants (e.g. 'Do prejudices about this group of immigrants exist in the Netherlands?'). We didn't develop an assignment covering acting democratically, because none of the participating classes signed up for that option.

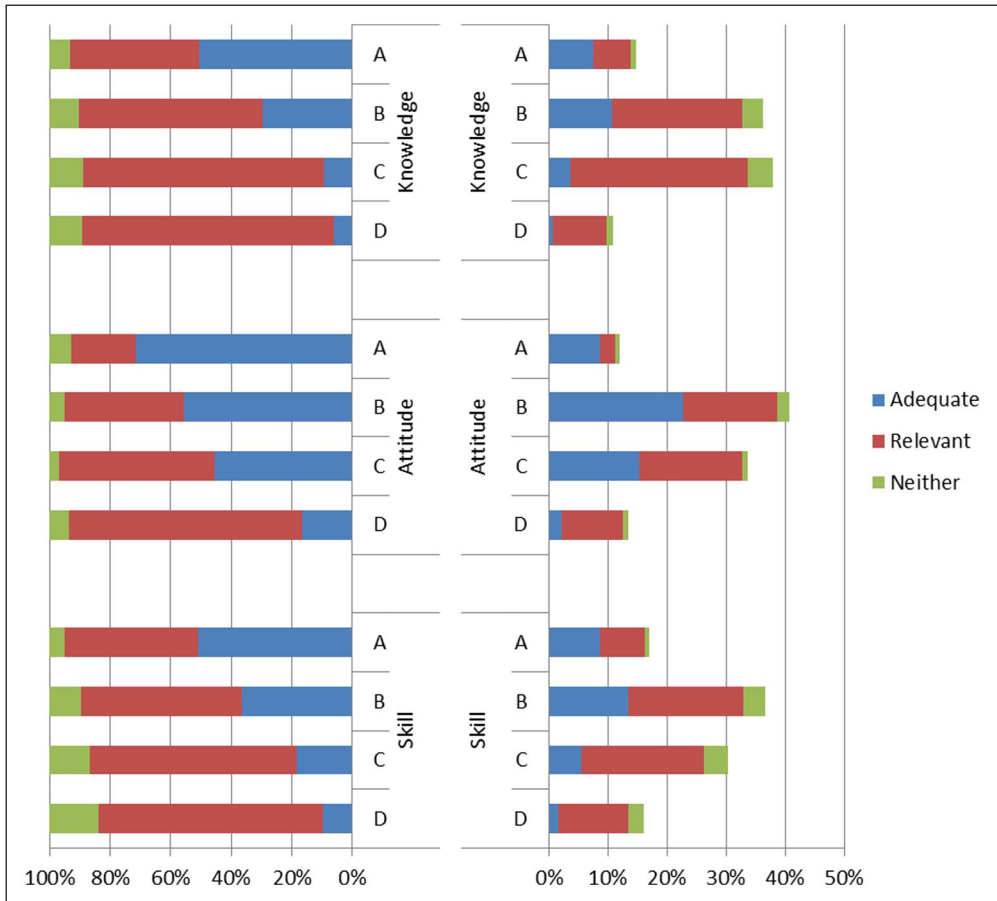
In both cases, students' answers were graded A, B, C, or D on the knowledge and attitude dimensions of the corresponding rubric. The responses of the first class completing the task (21 students), were graded by all four raters, who subsequently discussed any differences in grading. This discussion occurred after the student raters had been trained to use the rubrics and was combined with discussing the first 40 self-evaluations. The remaining student materials were assessed by three raters: by the first author and by two of the other raters in rotation. If at least two raters agreed about the student's level, the student was assigned that level. If there was no agreement at all, no level was assigned. Again, inter-rater agreement and Cohen's kappa were calculated to determine the reliability of scoring.

## **Results**

### *The use of rubrics for self-evaluation of citizenship competences*

To evaluate our instrument for self-evaluation of citizenship competences, we first investigated whether students are able to evaluate their own competences. In other words, are they able to assign their citizenship competences to one of the four levels of knowledge, attitudes and skills? The results show that 97% to 100% of students chose a level to indicate their 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% of students failed to make a clear choice. Figure 2 shows that most students evaluated 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 provided was relevant to the dimension (knowledge, attitude and skill) and the social task. For an explanation to be 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.



**Figure 2.** Results of student self-evaluation.

The graph on the left shows the percentage of adequate (and thus also 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 percentages are somewhat lower for skills, which shows that the students had slightly more difficulty providing a relevant explanation for their self-evaluated level of competence. The low percentage of relevant explanations for knowledge about acting socially responsible shows a

**Table 3.** Interrater agreement between assessors for self-evaluation.

|                         | 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    |

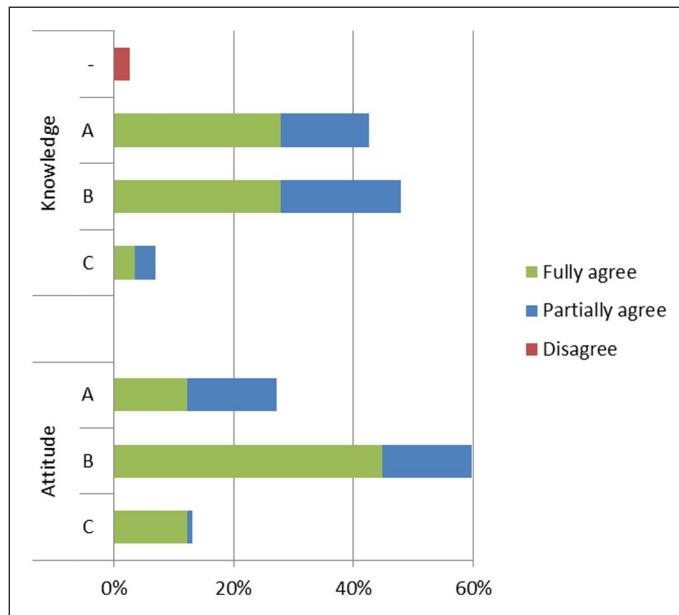
deviation from the pattern and appeared to be caused by students explaining their attitude rather than their knowledge.

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 rubric. 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 highly related to the proficiency level selected. Two reasons for not providing an adequate explanation stood out. First, students overestimated themselves, resulting in explanations that were more adequate for a lower level; secondly, students stayed too close to the text in the rubric by copying parts of it, so that their explanation, although relevant, could not be considered adequate.

Students' responses to each of the prompts were typically one or two sentences long. Students' explanations typically stayed very close to the descriptions in the rubric, particularly in response to each of the first prompts. They generally repeated part of the rubric, sometimes adding a brief explanation (e.g. 'I know why people's background can be important to them, and that some people find this more important than others'). The prompt to give an example appeared to help students provide a more personal explanation (e.g. 'A girl from our class last year got really upset when she considered something to be discriminating, while another did not seem to care at all'). However, a considerable number of students provided only little personal input, which often meant an explanation would not be considered adequate because the student had not provided sufficiently convincing evidence supporting the claim that a certain level fitted with their proficiency.

The results with regards to whether explanations were relevant and adequate are based on majority agreement between the raters concerned. Table 3 shows the agreement between the raters in this study. The average pairwise agreement shows the likelihood that any two raters agree. Agreement between pairs of raters was around 90% for relevance and around 80% for adequacy. Agreement between all three or four raters (i.e. full agreement) was 85% to 89% for relevance and 65% to 76% for adequacy. Cohen's kappa ranged from 0.164 to 0.838, with a mean of 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 magnitude of Cohen's kappa are essentially arbitrary, 0.4–0.6 can be considered to indicate moderate agreement (Landis and Koch, 1977). All mean Cohen's kappas are within this range.

Overall, the results of the self-evaluation show that nearly all students were able to assign themselves to a level, and nine out of 10 were able to provide a relevant explanation. Providing an adequate explanation for the level selected, however, appeared considerably more difficult. About half the students were able to provide an adequate explanation for their attitude but only one in four for knowledge and skills. As might be expected, the higher the level selected, the more difficult it was to provide an adequate explanation. Inter-rater agreement as expressed by Cohen's kappa was



**Figure 3.** Results of the reflection assignment.

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

generally moderate, indicating a moderate interrater agreement on how to evaluate the students' explanations. We believe this is acceptable because – given that the main purpose of the rubric approach is to provide room for personal interpretations – evaluating citizenship competences can be considered 'low-stakes' in the current context.

### *The use of rubrics for evaluating student citizenship competences*

Based on their answers to the reflection assignment task, all students could be assigned to a level for attitude and 97% for knowledge. Figure 3 shows the levels to which students were assigned and the extent to which the raters agreed. Most students scored A or B on knowledge and B on attitude.

The responses students gave to the questions in the assignment showed more variation in length than the self-evaluation task. Students were asked a set of questions and typically answered each question in either one or multiple sentences, with most students providing single sentence answers. Shorter answers typically still contained personal input, for example, 'I found it pretty confronting to see that some people have very little opportunity and that they are really low on the social ladder'. Longer answers contained some more elaborate reflection:

Because of this assignment you learn to see that some people are higher on the social ladder than others. I think the differences are pretty large, especially because everyone's always saying that we are all equal. Two students were on completely different ends of the social ladder. That should not be the case!

Table 4 shows the extent of agreement between raters on the students' level of proficiency. Agreement between pairs of raters was around 70% for knowledge and around 80% for attitude.

**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    |

In 59% of cases for knowledge and 70% of cases for attitude, all raters were in agreement. 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 kappas of 0.4–0.8 indicate moderate to substantial agreement (Landis and Koch, 1977).

The results of the reflection assignment show that the answers of nearly all students could be assigned to a level based on simple majority agreement between the raters. The overall agreement between raters can partly be explained by chance; however, after controlling for chance, the interrater 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 adequate explanations on the self-assessment assignment. The rubrics appear to be a reasonably reliable way to assess student knowledge and attitudes on the basis of their answers to an assignment with open-ended questions. The inter-rater reliabilities for the reflection assignment are slightly higher than for self-assessment, although the differences between the reliability coefficients are small.

## Discussion

In this study, we aimed to develop rubrics that provide a novel approach to the assessment of citizenship competences and examine their advantages and disadvantages for both self-evaluation and evaluation by assessors. Frequently used instruments such as tests and questionnaires are less suitable to reflect the wide range of personally relevant contexts in which student citizenship competences are applied (Daas et al., 2016). They do not provide an understanding of students' thinking about citizenship. Rubrics, on the other hand, can be used to facilitate an open-ended approach to assessing students' citizenship competences. We developed and tested three rubrics relating to aspects of citizenship relevant to the daily lives of young people. While rubrics are considered suitable for assessing complex competences (Jonsson and Svingby, 2007), no earlier studies have been conducted into their application for assessing citizenship competences.

Our first research question concerns the use of rubrics for *self-evaluation*. Our study shows that rubrics are a suitable tool for students to determine their own level of citizenship competences. The vast majority of students were able to assign their citizenship competences to one of the levels and to provide a relevant explanation for their self-evaluation. Students appeared to understand the topics addressed in the rubrics and could assess their own competences. They were generally able to elaborate on the relevance of the social task for their own lives, which allowed us to determine whether they could indeed give convincing arguments for their self-assessed level. It should however be noted, that these explanations often stayed very close to the contents of the rubric, and were thus of lower quality than we had hoped. For attitudes, 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. The majority of students selected level B or C. However, after checking their explanations, most students who provided adequate explanations placed themselves at level A or B. Students appeared to have more difficulty explaining what they know or are able to do with regards to a social task, than explaining what they think and feel.

The prompt to provide an example did appear to help students provide more personally relevant input. Students' explanations were often only one or two sentences long, which again is reflected in the low percentages of adequate explanations. With regards to the reliability of scoring, inter-rater agreement was found to be moderate.

Our second research question concerns the use of rubrics for *assessment of student work*. The rubrics allowed assessors to assign students to a level of knowledge and attitudes based on their answers to a set of open-ended questions in which they were asked to reflect on a civic issue in an everyday context. The results show that 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 the students' answers, with students who scored higher (C or D) often providing more elaborate answers. The input provided by students was generally more elaborate than in the self-evaluation task. Students again typically answered each question in one sentence, but the assignments contained a set of questions eliciting different responses. Some students provided answers of multiple sentences in which they explained their views and understanding. With regards to the reliability of scoring, inter-rater agreement was found to be moderate to substantial.

Our main research question – *What are the advantages and disadvantages of using rubrics to assess students' citizenship competences?* – can now be answered. The main advantage of using rubrics is the room they offer for students to provide personally meaningful context and experiences from their own lives. With the inherent diversity of viewpoints that are part and parcel to citizenship, using rubrics aligns with the multidimensionality of the concept. However, for students to provide these input as well as rating these appear to be easier for attitudes than for knowledge and skills. Rating students' responses suggested students overestimate their proficiency. This means that for teachers, one of the advantages of using rubrics seems to be that asking students to explain their views can help identify students' misconceptions, which is generally not possible with the Likert-type items normally found in tests and questionnaires.

There are several disadvantages to the use of rubrics. Inter-rater reliability was moderate to substantial in our study. Like for many other rubrics, this can be considered sufficient for low-stakes assessment, but not for high-stakes summative assessment (Jonsson and Svingby, 2007). Another common practical disadvantage of rubrics (compared to tests and questionnaires), is the time it takes for both students and teachers to administer the instrument and assess students' answers. Finally, both approaches to assessing students' citizenship competences applied in this study relied on students' written responses. Although there were definitely a considerable amount of exceptions, most students only provided short answers, which severely limited the amount of information (and thus context) they shared.

The use of rubrics for assessing student citizenship competences thereby raises the question of the influence of their reflection and writing skills. On the self-evaluation task, half of the students provided an adequate explanation for their attitudes and one in four students for their knowledge and skills. The large number of inadequate explanations appeared to be due to two reasons: (1) students overestimating their own competences and (2) inadequate ability or motivation to provide grounds for their selected level. While it is impossible to distinguish between these reasons based on the data, the results may be improved if students are encouraged to provide better explanations or given support to do so. Several teachers reported that they needed to explain the rubrics to students, which shows that the students in our study were probably unfamiliar with this type of task. Open-ended approaches such as rubrics make greater demands on students' reflection skills, which makes teacher guidance even more important. The cognitive load which the assignment placed on their writing skills may have played a role too. The quality of student writing has been found to influence assessor judgements (Rezaei and Lovorn, 2010). Finally, motivation may have played a

role. This was a low-stakes assessment and this may have had a negative impact on students' motivation to make an effort (cf. Richardson, 2010).

In light of the moderate to substantial agreement between raters, it is unclear if agreement between raters would increase given more time and training or if they are the result of differences of opinions between raters about the meaning of the cells making up the rubrics. In our study, raters were trained by providing explanations of the rubrics before they used them, and the assessments of the first students were discussed with the raters. We noticed hardly any increase in agreement after they had discussed the initial results. To reach higher inter-rater reliability, the raters could perhaps have collaborated more intensively (cf. Meier et al., 2006). Jonsson and Svingby (2007) note that reliability is self-evidently lower when students are given more freedom to complete a performance task. Since one of the aims of assessing citizenship competences through rubrics is to provide students with room for personally meaningful interpretations of citizenship, we consider the level of agreement acceptable for assessing citizenship competences for low-stakes purposes.

In conclusion, our results show that an open-ended approach to assessing citizenship competences based on rubrics could be a useful addition to evaluations using tests and questionnaires. Using rubrics provides room for students to describe what citizenship means to them on the basis of their own experiences. Asking students to share personally meaningful experiences and viewpoints might make them more aware of their own experiences and views on citizenship (cf. Vaessen et al., 2022). For teachers, one of the benefits of using rubrics is that asking students to explain their views may enable specific interventions in the learning process.

The linking of citizenship competences to contexts related to the students' lives increases the validity of the measurement and potentially make citizenship and citizenship knowledge, skills and attitudes more meaningful for students. Furthermore, this contextualization increases opportunities for specific reflection on students' knowledge, attitudes and skills by both students themselves and teachers. This may thus contribute to promoting deliberation of the concept of citizenship, its relevance *and* opportunities for shaping the learning process. Further development of using rubrics to assess citizenship competences might particularly benefit from approaches that stimulate students to share more information on their views and experiences to further increase the reliability and validity of such an assessment. The rubrics we developed for this study were designed to reflect citizenship of young people in the Dutch context. Although the Dutch context is not substantially different from that of other Western countries, we would welcome further investigation of the application of rubrics for citizenship in other contexts.

### **Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was financially supported by the Dutch Inspectorate of Education.

### **ORCID iDs**

Remmert Daas  <https://orcid.org/0000-0002-7348-7126>

Geert ten Dam  <https://orcid.org/0000-0003-2929-7920>

### **Supplemental material**

Supplemental material for this article is available online.

## References

- Abowitz KK and Harnish J (2006) Contemporary discourses of citizenship. *Review of Educational Research* 76(4): 653–690.
- Andrade HG (2005) Teaching with rubrics: The good, the bad, and the ugly. *College Teaching* 53(1): 27–31.
- Brookhart SM (2018) Appropriate criteria: Key to effective rubrics. *Frontiers in Education* 3(22): 1–12.
- Brookhart SM and Chen F (2015) The quality and effectiveness of descriptive rubrics. *Educational Review* 67(3): 343–368.
- Cogan JJ and Morris P (2001) The development of civics values: An overview. *International Journal of Educational Research* 35: 1–9.
- Council of Europe (2016) *Competences for Democratic Culture: Living Together as Equals in Culturally Diverse Democratic Societies*. Strassbourg: Council of Europe Publishing.
- Daas R, Ten Dam G and Dijkstra AB (2016) Contemplating modes of assessing citizenship competences. *Studies in Educational Evaluation* 51: 88–95.
- Dawson P (2017) Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education* 42(3): 347–360.
- De Groot I, Daas R and Nieuwelink H (2022) Education for democratic citizenship in Dutch schools: A bumpy road. *Journal of Social Science Education* 21(4): 1–31.
- Eidhof BB, Ten Dam GT, Dijkstra AB et al. (2016) Consensus and contested citizenship education goals in Western Europe. *Education Citizenship and Social Justice* 11(2): 114–129.
- Eurydice (2017) *Citizenship Education at School in Europe*. Luxembourg: Publications Office of the European Union.
- Geboers E, Geijsel F, Admiraal W et al. (2015) Citizenship development of adolescents during the lower grades of secondary education. *Journal of Adolescence* 45(1): 89–97.
- Hoskins B, Villalba C and Saisana M (2012) *The 2011 Civic Competence Composite Indicator (CCCI-2)*. Luxembourg: Publications Office of the European Union.
- Inspectorate of Education (2016) *Burgerschap op school: Een beschrijving van burgerschapsonderwijs en de maatschappelijke stage* [Citizenship in schools: A description of citizenship education and community service].
- Jonsson A and Svingby G (2007) The use of scoring rubrics: Reliability, validity and educational consequences. *Review of Educational Research* 2(2): 130–144.
- Keating A, Kerr D, Benton T et al. (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: Department for Education.
- Landis JR and Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1): 159–174.
- Lawy R and Biesta G (2006) Citizenship-as-practice: The educational implications of an inclusive and relational understanding of citizenship. *British Journal of Educational Studies* 54(1): 34–50.
- Meier SL, Rich BS and Cady J (2006) Teachers' use of rubrics to score non-traditional tasks: Factors related to discrepancies in scoring. *Assessment in Education Principles Policy and Practice* 13(1): 69–95.
- Nieuwelink H and Oostdam R (2021) Time for citizenship in teacher training. *Journal of Social Science Education* 20(3): 130–146.
- Nieuwelink H, Ten Dam G, Geijsel F et al. (2018) Growing into politics? The development of adolescents' views on democracy over time. *Politics* 38(4): 395–410.
- Oser F and Veugelers W (2008) *Getting involved: Global citizenship development and sources of moral values*. Rotterdam: Sense Publishers.
- Osler A and Starkey H (2006) Education for democratic citizenship: A review of research, policy and practice 1995–2005. *Research Papers in Education* 21(4): 433–466.
- Panadero E and Jonsson A (2013) The use of scoring rubrics for formative assessment purposes revisited: A review. *Review of Educational Research* 9: 129–144.
- Pancorbo G, Primi R, John OP et al. (2021) Formative assessment of social-emotional skills using rubrics: A review of knowns and unknowns. *Frontiers in Education* 6(6): 687661.



- Reddy YM and Andrade H (2010) A review of rubric use in higher education. *Assessment & Evaluation in Higher Education* 35(4): 435–448.
- Rezaei AR and Lovorn M (2010) Reliability and validity of rubrics for assessment through writing. *Assessing Writing* 15(1): 18–39.
- Richardson M (2010) Assessing the assessment of citizenship. *Research Papers in Education* 25(4): 457–478.
- Rychen DS and Salganik LH (2003) *Key Competencies for a Successful Life and Well-Functioning Society*. Cambridge: Hogrefe & Huber Publishers.
- Schulz W, Ainley J, Fraillon J et al. (2016) *International civic and citizenship education study 2016: Assessment framework*. Amsterdam: IEA.
- Schulz W, Fraillon J, Ainley J et al. (2008) *International Civic and Citizenship Education Study: Assessment Framework*. Amsterdam: IEA.
- Ten Dam G, Geijsel F, Reumerman R et al. (2011) Measuring young people's citizenship competences. *European Journal of Education* 46(3): 354–372.
- Ten Dam G and Volman M (2007) Educating for adulthood or for citizenship: Social competence as an educational goal. *European Journal of Education* 42(2): 281–298.
- Tierney R and Simon M (2004) What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research and Evaluation* 9(2): 1–7.
- Torney-Purta J, Lehmann R, Oswald H et al. (2001) *Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen*. Amsterdam: IEA.
- United Nations (2015) *Transforming Our World: The 2030 Agenda for Sustainable Development (A/RES/70/1)*. New York, NY: United Nations.
- Vaessen A, Daas R and Nieuwelink H (2022) All things considered: The views of adolescents in vocational education on competing democratic values. *Citizenship Social and Economics Education* 21(1): 3–21.
- Westheimer J and Kahne J (2004) What kind of citizen? The politics of educating for democracy. *American Educational Research Journal* 41(2): 237–269.