

Research Article

Copyright © RIGEO 2016

**To cite this article:** Bijsterbosch, H.; Van Der Schee, J.; Kuiper, W.; Béneker, T. (2016). "Geography Teachers' Practices regarding Summative Assessment: A Study of Pre-Vocational Education in the Netherlands", *RIGEO*, 6 (2), 118-134, Retrieved from <http://www.rigeo.org/vol6no2/Number2Summer/RIGEO-V6-N2-1.pdf>

**Submitted:** April 18, 2016

**Revised:** August 01, 2016

**Accepted:** August 18, 2016

## Geography Teachers' Practices regarding Summative Assessment: A Study of Pre-Vocational Education in the Netherlands

Erik BIJSTERBOSCH<sup>1</sup>

*Windesheim University of Applied Sciences, Zwolle, The Netherlands*

Joop VAN DER SCHEE<sup>2</sup>

*VU University, Amsterdam, The Netherlands*

Wilmad KUIPER<sup>3</sup>

*Utrecht University, Utrecht, The Netherlands*

Tine BÉNEKER<sup>4</sup>

*Utrecht University, Utrecht, The Netherlands*

### Abstract

To start a teacher professional development programme on the relationship between classroom summative assessment and learning, the current practices and dispositions of geography teachers towards internal school-based examinations in pre-vocational education in the Netherlands were investigated. A questionnaire provided data on how teachers construct these examinations and how they perceive the extent to which they use test items in these examinations that appeal to distinct cognitive processes. The data were statistically analysed to explore teachers' practices regarding the construction of the examinations and the correlation with their perceptions on test items appealing to distinct cognitive processes. The results showed that teachers rarely construct test items themselves; instead, they rely to a considerable degree on test items created by outside sources. In particular, older teachers and teachers with greater teaching experience tend

<sup>1</sup> **Corresponding Author:** Ph. D. Candidate; Windesheim University of Applied Sciences, Department of Education, Campus 2-6, 8017 CA Zwolle, The Netherlands, 0031-88-4696352, H.Bijsterbosch[at]windesheim.nl

<sup>2</sup> Prof. Dr., Faculty of Earth and Life Sciences, VU University, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands, J.A.vander.Schee[at]vu.nl

<sup>3</sup> Prof. Dr.; Utrecht University, Freudenthal Institute, Princetonplein 5, 3584 CC Utrecht, The Netherlands, W.Kuiper[at]uu.nl

<sup>4</sup> Assoc. Prof. Dr.; Utrecht University, Faculty of Geosciences, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands, T.Beneker[at]uu.nl

to use more test items from outside sources. According to the respondents, about two-thirds of the test items appeal to higher cognitive processes. When teachers do construct test items themselves, however, they perceive to use more test items that appeal to higher cognitive processes. Furthermore, teachers' dispositions regarding the purpose of the internal school-based examinations seem to be highly influenced by high-stakes tests, such as the national exam.

### Keywords

Geography Education, Teacher Professional Development, Teachers' Practices, Teachers' Dispositions, Summative Assessment

---

The relationship between teachers' practices on classroom assessment and learning has been studied extensively in recent decades. Several reviews and studies on this relationship have been published (Black & Wiliam, 1998a; Black, Harrison, Hodgen, Marshall, & Serret, 2010, 2011; Harlen, 2004a, 2004b; Harlen & Deakin Crick, 2002, 2003). Some of the reviews or studies focused on the relationship between assessment and learning (Black et al., 2011; Black & Wiliam, 1998b); others focused on the relationship between summative assessment and teachers' practices (Black et al., 2010; Harlen, 2004b) or on the relationship between students' learning and motivation (Harlen & Deakin Crick, 2002).

A central issue in research on the relationship between assessment and learning is the effect of assessment—formative as well as summative—on meaningful learning. Meaningful learning is generally defined as a learning process in which learners actively construct knowledge by integrating new information with existing knowledge. This concept has also been equated with cognitive processes that transcend rote learning (Anderson & Krathwohl, 2001). Cognitive processes transcending rote learning include processes such as understanding, applying, analysing, evaluating and creating.

The relationship between assessment and meaningful learning is not always positive. In their review on classroom formative assessment, Black and Wiliam (1998a) revealed that classroom evaluation practices tend to encourage rote learning instead of meaningful learning. Others emphasized a similar tendency of teachers to focus classroom summative assessments on fact recall and to a lesser extent on critical thinking or other complex and demanding cognitive skills (Harlen, 2005; James & Gipps, 1998).

The tendency of teachers to use test items in classroom summative tests that focus on recall and memorization is often strengthened by the impact of high-stakes tests (Klenowski & Wyatt-Smith, 2011). In high-stakes tests such as national exams, test items in general put more emphasis on rote learning instead of meaningful learning (Harlen, 2005; James & Gipps, 1998). This tendency is enforced by the requirement to use test formats that produce reliable test results (Harlen, 2005; Stimpson, 2006). Reliable test results are best achieved using test items that can be readily and reliably marked (Harlen, 2005). Teachers frequently imitate these formats for test items in their classroom summative assessments (Black & Wiliam, 1998a) and train their pupils how to answer these specific test items (Harlen, 2005). This phenomenon is also referred to as “teaching to the test.”

The same practices can also be found in geography education. A study on assessment practices in K-12 classrooms and large-scale assessments in the USA revealed that a majority of the assessments test students' recall of geographic facts (Wertheim, Edelson, & The Road Map Project Assessment, 2013), while less than a third of the test items appealed to higher cognitive processes and the application of geographical skills. The tendency to give priority to test items that seem to ensure reliable test scores has been observed by others as well (Stimpson, 2006; Weeden, 2013)

The impact of high-stakes tests as such undermines teachers' attention to validity issues in summative assessments (Black et al., 2010). Teachers seem to be more concerned with issues of reliability and accountability than whether the assessments are in line with subject specific objectives. To achieve that teachers will pay more attention to the issue of validity, Black et al. (2010, p227) suggest 'that an appeal to the belief and values that underlie their commitment to their subjects can be a way to make validity a more salient feature of their work'. Besides reflection on their beliefs and values, attention should be given to teachers' assessment skills.

Teachers' practices in classroom summative assessment are not only influenced by high-stakes tests but also by teachers' dispositions regarding summative assessment. Teachers' dispositions can be described as the ability of teachers to apply knowledge and skills attuned to their beliefs and values. Jo and Bednarz (2014, p. 199) defined teachers' dispositions as 'the tendencies of a teacher's behaviour employing particular knowledge and skills to achieve certain teaching goals'.

Teachers' dispositions are the interplay between three separate yet interconnected domains; the intellectual, the cultural and the moral domain (Schussler, 2006; Schussler, Stooksberry, & Bercaw, 2010). These three domains reflect the subject specific content, the identities of teachers and their values. For teachers it is not sufficient to have the knowledge and skills when they are not willing to employ or enact them in their classroom (Jo & Bednarz, 2014). The will to enact pedagogical content knowledge in the classroom is affected by teachers' personal beliefs and values.

Teachers' tendency to use test items in classroom summative tests that focus on recall and memorization are influenced, therefore, by teachers' dispositions towards summative assessment as well. Teachers, however, are not always aware of the fact that validity is under pressure because of this tendency. The review by Black and Wiliam (1998a) revealed that there seems to be a lack of consistency between the teachers' classroom practices and their perceptions of learning. Teachers are often unaware that they focus on rote learning. According to the authors, teachers often emphasize that they want to develop understanding as part of meaningful learning with their students.

Another important finding of previous research on teachers' practices regarding summative assessment is that teachers rarely discuss or share their practices with colleagues in the same school (Black & Wiliam, 1998b). Teachers are not only unaware of their colleagues' practices but do not trust assessment results obtained from their colleagues either. However, working together with their colleagues can improve teachers' practices. Harlen (2005) showed a promising effect on teachers' practices when they share their understanding of assessment procedures.

---

A lack of professional collaboration on assessment practices was also noted in previous research from the USA. Cizek, Fitzgerald, & Rachor (1996) showed a substantial variation in assessment practices between teachers. First, the use of primary sources to develop minor and major assessments varies between teachers. For major tests, teachers rely more on test materials from private publishers than for minor tests. However, not all teachers rely on test items from outside sources to the same degree. For instance, the use of test items from private publishers is influenced by characteristics of a teachers' work experience. Beginning teachers used fewer test items from outside sources and developed more minor tests themselves than did more experienced teachers.

### **Context: Geography Education in the Netherlands**

This paper is one part of a research effort designed to explore how summative assessment in pre-vocational geography education in the Netherlands—and more specifically, internal school-based examinations—can contribute to meaningful learning. Furthermore, this research design explores how geography teachers can be trained and scaffolded to construct, judge and mark test items for internal school-based examinations in pre-vocational geography education in the Netherlands that contribute to meaningful learning.

The examination programme in pre-vocational education in the Netherlands consists of two parts; the first part pertains to about two-thirds of the objectives of the examination programme and is assessed summatively with internal school-based examinations, while the second part pertains to the other one-third of the objectives and is assessed summatively with a national external end-of-school (exit) examination. Both parts, however, contribute equally (50% each) to the overall result.

A previous study at the first stage of the research design provided some insight of how the objectives for the internal school-based examinations are assessed (Bijsterbosch, van der Schee & Kuiper, 2016). This study was based on a content analysis of internal school-based examinations in pre-vocational education in the Netherlands. The study revealed that the test items on the internal school-based examinations in pre-vocational education in the Netherlands tend to focus on rote learning. Over 60% percent of the test items appeal to some type of remembering, almost 30% to understanding and only 10% to higher cognitive processes such as applying, evaluating and creating. Most test items reflected the formats for test items that were used in former national external end-of-school (exit) examinations. The outcomes of this first study are in line with previously mentioned outcomes from the literature; that is, teachers tend to focus their classroom assessments on rote learning.

Still, little is known about the practices, beliefs and values of geography teachers regarding the construction of their internal school-based examinations in the Netherlands. This study attempts to provide some insight into what the current practices and dispositions of geography teachers in pre-vocational education in the Netherlands regarding internal school-based examinations are. The main issues with respect to teachers' practices that are investigated here are to what extent do teachers construct test items themselves, whether they work collaboratively on the examinations and if and how they use instruments such as taxonomies or test matrices.

This study will also provide some insight concerning teachers' dispositions regarding the purpose of these internal school-based examinations and the relationship with the external end-of-school (exit) examination. Moreover, teachers' perceptions of the extent to which they use test items that appeal to higher cognitive processes will be explored. This perception is closely related to one of the three domains of teachers' dispositions: the intellectual domain. Intellectual dispositions have been defined by Schussler, Stooksberry & Bercaw (2010, p. 352) as 'teachers' inclination to process knowledge of content and pedagogy, their awareness of what the educational context requires for desired learning outcomes to be reached, and their inclination to put their knowledge and awareness to use accordingly in the classroom'. The outcomes on teachers' perceptions of the extent to which they use test items that appeal to higher cognitive processes should provide some insight into teachers' intellectual dispositions. The relationship between teachers' perceptions and practices towards test items that appeal to higher cognitive processes will be explored as well.

Finally, the relationships between teachers' practices and some background characteristics will be investigated to explore whether teachers' practices are influenced by age or teaching experience. The research questions for this study are as follows:

1. What are the current practices, beliefs and values of geography teachers in pre-vocational secondary education in the Netherlands regarding internal school-based examinations?
2. What is the relationship between geography teachers' practices in pre-vocational secondary education in the Netherlands and their perceptions of test items that appeal to distinct cognitive processes in their internal school-based examinations?
3. What is the relationship between the background characteristics of geography teachers in pre-vocational secondary education in the Netherlands and their practices regarding the construction of school-based examinations?

## **Methodology**

This study analysed the responses to a questionnaire that included 21 questions as well as some items concerning the background characteristics of the respondents (such as their age and teaching experience). The questionnaire was divided into four parts.

In the first part, respondents were asked to provide information about the content of the internal school-based examinations in relation to the objectives of the examination programme. This part addresses the content validity of the internal school-based examinations. The second part contained questions about the application of instruments such as test matrices or taxonomies. We also asked the teachers whether they constructed test items themselves or collaboratively and whether they used test items from outside sources such as textbooks. The respondents were also asked to indicate the percentage of test items in their school-based examinations that are self-constructed or come from outside sources. The third part was about the teachers' perceptions with regard to distinct cognitive processes in the internal school-based examinations. The respondents were

---

asked to give an indication of the percentage of test items in their school-based examinations that appeal to distinct cognitive processes. The final part contained questions about the beliefs and conceptions of the teachers concerning the objectives and goals of internal school-based examinations in the examination programme.

The questionnaire was piloted in 2014 on 4 geography teachers and adjusted based on their feedback. Subsequently, an invitation to fill in the questionnaire was sent to geography teachers who worked in pre-vocational education. The questionnaire was published online, and a letter with a hyperlink was distributed by networks of secondary education teacher training institutions, by the newsletter of the Royal Dutch Geographical Society (KNAG) and by a national online community of geography teachers. The data were collected in October and November 2014.

Out of a total number of 729 schools offering the theoretical programme of pre-vocational secondary education in the Netherlands where geography was one of the possible subjects, 74 respondents filled out the online questionnaire. These figures roughly indicate that approximately 10% of the teachers with a group of pupils in the examination programme responded to the questionnaire.

Questionnaire responses were first analysed on a descriptive level. Then, several tests were performed to explore the correlations between variables (Pearson's PMCC, Spearman's RCC, Chi-Square and Cramer's V) and to explore differences in the means of variables (t-tests, ANOVA). Correlations and differences in means were regarded as significant when  $\alpha < 0,05$ .

### **Background Characteristics of Respondents**

Of the 74 respondents, 45 were male, and 29 were female. All the respondents worked as geography teachers in the theoretical programme of pre-vocational secondary education (VMBO-gt) in the Netherlands. The mean age of the respondents was 40 ( $SD = 10.95$ ). The mean number of years of teaching experience was 13 ( $SD = 9.06$ ). Approximately 81% of the respondents had a bachelor's degree in geography education, and 14% had a master's degree in geography education. The other respondents either had a bachelor's degree in primary education or held no valid qualification (yet).

## **Findings**

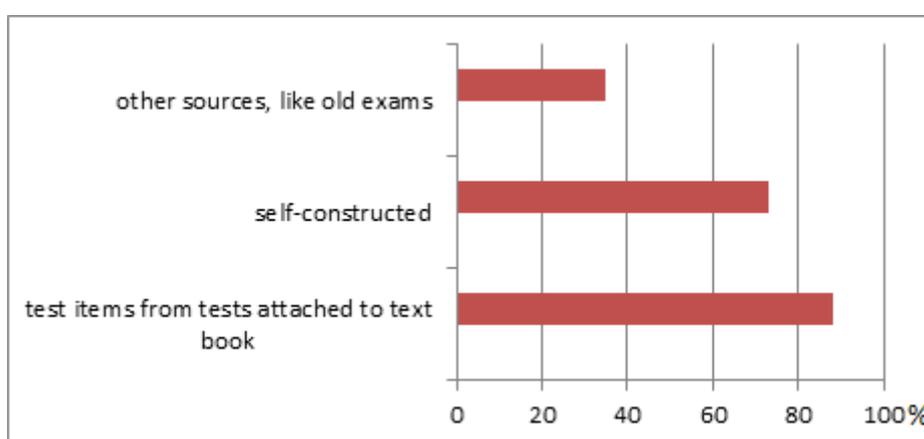
This section presents the results for the three research questions. The first part of this section reports the findings concerning geography teachers' practices with respect to the construction of internal school-based examinations in pre-vocational education in the Netherlands and their beliefs and values. The second part reports the findings concerning the relationship between these practices and their perceived appeal to distinct cognitive processes. The third and final part reports the findings concerning the relationship between teachers' practices and their background characteristics.

### **Current Practices, Beliefs and Values of Geography Teachers**

First, the current practices of geography teachers regarding the construction of school-based examinations were investigated. Respondents were asked to answer questions about the origin of the test items they use and the perceived percentage of test items from

different sources. Second, the respondents were asked which instruments they use to determine the content of the internal school-based examinations. Finally, the respondents were asked whether they work collaboratively on the construction of internal school-based examinations.

To the question about the origin of test items in the internal school-based examinations, most teachers responded that they use multiple sources. The respondents could choose between tests attached to the textbook, self-constructed test items or other sources such as older exams. The results show that respondents use more than one source to compose the school-based examinations. Most respondents use test items from tests attached to the textbook (88%) and also self-constructed test items (73%) (Figure 1).



*Figure 1.* Percentage of teachers that use test items from different sources

The respondents were also asked to give an indication of the percentage of test items in their school-based examinations related to various sources. Teachers responded that they perceive that almost 45% of the test items in the school-based examinations come from the tests attached to the textbooks and that only 17% are self-constructed test items (Table 1).

To know how geography teachers determine what the content of the internal school-based examinations will be, the questionnaire contained some questions about whether teachers use instruments to determine the content of the examinations, especially a taxonomy or test matrix.

The responses indicate that 78% use some type of taxonomy to construct internal school-based examinations. More than half the respondents use a taxonomy that has become well known in the Netherlands, the so called RTTI taxonomy that consists of four categories: remembering (R), executing a familiar task (T1), implementing an unfamiliar task (T2) and comprehension (I). The other teachers use taxonomies such as Bloom's taxonomy. Approximately 22% of the teachers use no taxonomy.

Table 1  
*Perceived Percentage of Test Items Related To the Origin of Test Items*

	% test items from textbook	% test items self-constructed	% test items from older exams	% other sources
Mean	45	17	29	5*
SD	30,3	21,5	10,7	8,6

*\*Note: the sum of the perceived percentages is 96 because the respondents had to fill out an estimated percentage for the various categories and the sum of the estimation of some respondents was less than 100 percent.*

Responses regarding the application of a test matrix (Table 2) show that 38% of the teachers do not use a test matrix to construct the internal school-based examinations. This percentage is slightly higher than the percentage of teachers who do not use a taxonomy, which indicates that some teachers use a taxonomy without a test matrix. The teachers who do use a test matrix rely on the matrix from the instructor's textbook (28%), construct a test matrix by themselves (18%) or construct a matrix in collaboration with their colleagues (10%)

When asked whether teachers work collaboratively on the construction of the school-based examinations, 65% of the respondents reported that they collaboratively decide what subject specific content in relation to the objectives of the examination programme will be assessed in the internal school-based examinations. Furthermore, 52% of the teachers work collaboratively on the construction of the test items; about two-thirds of the respondents decide in collaboration with their colleagues what the caesura of the internal school-based examinations will be, and 40% work collaboratively on the correction of the school-based examinations.

Table 2  
*Percentage of Teachers Who Use a Test Matrix When Constructing School-Based Internal Examinations*

Test matrix	Percentage
No test matrix	38%
Textbook matrix	28%
Self-made matrix	18%
Matrix made in collaboration with colleagues	10%
Other	2%
No response on the question	4%

### **Teachers' Beliefs and Values**

To investigate geography teachers' beliefs and values regarding the internal school-based examinations the respondents were asked what the content of the internal school-based examinations should be and what they thought the purpose of the internal school-based examinations should be. Almost all (90%) responded that the objectives of the examination programme that belong to the internal school-based examinations as well as

the final exam should be assessed in the internal school-based examinations. Furthermore, 74% of the teachers responded that one purpose of the internal school-based examinations is to prepare the pupils for the external end-of-school (exit) examination by using the same formats for test items in the internal school-based examinations as in the final exam.

To explore the perceptions of teachers as to what extent they use test items that appeal to distinct cognitive processes, the respondents were asked to give an indication of the percentage of test items on their internal school-based examinations related to one of the following cognitive processes: (1) remembering, (2) understanding, (3) applying, and (4) other higher cognitive processes such as evaluating and creating. The results are summarized in Table 3.

According to the respondents, the test items in the cognitive dimension were rather equally divided over three of the cognitive processes, with a mean of 34% (*SD* 13.6) of the test items referring to remembering, 32% (*SD* 9.5) to understanding and 26% (*SD* 11.4) to applying. The perceived percentage of test items related to higher cognitive processes such as evaluating and creating was 12% (*SD* 6.5). The range between the minimum and maximum scores is rather wide, especially for remembering, understanding and applying.

Table 3  
*Scores of Teachers on the Perceived Percentage of Assessed Cognitive Processes in Internal School-Based Examinations*

Cognitive processes	n	Minimum	Maximum	Mean	SD
<i>remembering</i>	74	10	75	34	13.6
<i>understanding</i>	74	10	70	32	9.5
<i>applying</i>	74	5	70	26	11.4
<i>other higher cognitive processes such as evaluating</i>	74	0	35	12	6.5

### **Relationship between Current Practices and Cognitive Processes**

To explore the relationship between teachers' practices and their perceptions as to the extent to which they use test items that appeal to distinct cognitive processes, several correlation tests were performed. A positive correlation was found between the perceived percentage of test items from tests attached to the textbook and the perceived percentage of test items related to remembering ( $r = 0.29$ ,  $p < 0.05$ , two tailed) and a negative correlation between the perceived percentage of test items from tests attached to the textbook and the perceived percentage of test items related to higher cognitive processes ( $r = -0.33$ ,  $p < 0.01$ , two tailed). A converse pattern was found for the correlation between the perceived percentage of self-constructed test items and remembering ( $r = -0.24$ ,  $p < 0.05$ , two tailed) and the perceived percentage of self-constructed test items and higher cognitive processes ( $r = 0.38$ ,  $p < 0.01$ , two tailed).

A t-test for differences in means between the group of respondents who construct test items themselves and the group of respondents who do not shows a significant difference

for the perceived percentage of test items related to higher cognitive processes ( $t(72) = -2,05$ ,  $p = 0,044$ ). The respondents who construct test items themselves perceive the percentage of test items related to higher cognitive processes as being almost twice as high ( $M = 9,6$ ,  $SD = 7,9$ ) as the respondents who do not ( $M = 5,5$ ,  $SD = 5,9$ ).

A t-test for differences in means between the group of respondents who use test items from the tests attached to the textbook and the group that does not gives a similar result with respect to the perceived percentage of test items related to higher cognitive processes. The respondents who use test items from the tests attached to the textbook perceive the percentage of test items related to higher cognitive processes as lower ( $M = 7,8$ ,  $SD = 7,3$ ) than those who do not ( $M = 17,5$ ,  $SD = 6,1$ ). The t-test for this difference in means is also statistically significant ( $t(72) = 3,18$ ,  $p = 0,002$ ).

No statistically significant correlations were found between the use of a test matrix or taxonomy on the one hand and the perceived percentage of test items related to distinct cognitive processes. Additionally, no correlation was found between the collaborative construction of school-based examinations and the perceived percentage of test items related to distinct cognitive processes, nor for the relationship between the collaborative construction of school-based examinations and the use of instruments such as a test matrix or taxonomy.

### **Relationship between Background Characteristics and Teachers' Practices**

Finally, the relationship between some background characteristics such as age and teaching experience and teachers' practices were investigated. To test these relationships, the respondents were clustered into four categories by age and teaching experience. Correlation tests were run to test the relationship between age and teaching experience on the one hand and the origin of test items on the other. The results show a slight positive correlation for age and the number of respondents who use test items from tests attached to the textbook ( $r_s = 0,24$ ,  $p < 0,05$ , two tailed) and a slight negative correlation between age and respondents who construct test items themselves ( $r_s = -0,24$ ,  $p < 0,05$ , two tailed).

A correlation was also found between teaching experience and the perceived percentage of test items related to the different sources. The results show a slight positive correlation for teaching experience and the perceived percentage of test items from tests attached to the textbook ( $r_s = 0,25$ ,  $p < 0,05$ , two tailed). The negative correlation between teaching experience and the perceived percentage of self-constructed test items is somewhat stronger ( $r_s = -0,32$ ,  $p < 0,05$ , two tailed).

To explore in more detail the differences for age and the perceived percentage of self-constructed test items, an analysis of variance (ANOVA) was performed. The results show a significant difference by age class ( $F(3, 70) = 3,37$ ,  $p = .023$ ). Respondents between 22 and 30 years old perceive the percentage of self-constructed test items as higher than do respondents who are 51 years of age or older (Table 4). Post hoc tests show that this difference is statistically significant ( $p = .016$ ). The difference in perception of self-constructed test items can be explained by 13% for these two age groups ( $\eta^2 = .13$ ).

Table 4  
Age in Years and Perceived Percentage of Self-Constructed Test Items

Age	n	Mean	SD
22-30	20	27	27,7
31-40	19	20	21,0
41-50	19	15	18,7
51+	16	6	6,5
Total	74	17	21,5

An analysis of variance for teaching experience and the perceived percentage of test items from the tests attached to the textbook also shows a significant difference between these groups ( $F(3, 70) = 3,73, p = .015$ ). Teachers with 25 years or more of teaching experience perceive the percentage of test items from the tests attached to the textbook as higher than 69% (Table 5). Post hoc tests note a significant difference between the group of respondents with between 5 and 14 years of teaching experience and the group with 25 years or more of teaching experience ( $p = .016$ ). The difference between these groups can be explained by 14% for years of teaching experience ( $\eta^2 = .14$ ). The difference between the group with between 15 and 24 years of teaching experience and the group with 25 years or more of teaching experience is also significant. Post hoc tests for these two groups show a significant difference ( $p = .037$ ). The difference between these groups can be explained by 14% for years of teaching experience ( $\eta^2 = .14$ ).

Table 5  
Teaching Experience in Years and Perceived Percentage of Test Items from Tests Attached to the Textbook

Teaching experience in years	n	Mean	SD
0-4	12	51	28,1
5-14	38	38	29,2
15-24	13	36	30,1
25+	11	69	26,1
Total	74	45	30,3

The relationship between age and teaching experience and the perceived percentage of test items related to distinct cognitive processes was also explored. No significant correlation between age and the perceived percentage of test items related to distinct cognitive processes was found. Additionally, no statistically significant correlation between teaching experience and the perceived percentage of test items related to distinct cognitive processes was found.

Although no statistically significant correlation was found, an analysis of variance revealed that there is a significant difference between the classes on teaching experience and perceived percentage of test items related to higher cognitive processes ( $F(3, 70) = 3,47, p = .021$ ). Teachers with between 5 and 14 years of teaching experience perceive

the percentage of test items related to higher cognitive processes as higher (Table 6). Post hoc tests show that the difference between the groups with between 5 and 14 years of teaching experience and the group with 25 years or more of teaching experience on the perceived percentage of test items related to higher cognitive processes is significant ( $p = .034$ ). The difference between these groups can be explained for years of teaching experience by 13% ( $\eta^2 = .13$ ). An ANOVA test on age by class and the perceived percentage of test items related to higher cognitive processes showed no significant results.

Table 6  
*Teaching Experience in Years and Perceived Percentage of Higher Cognitive Processes*

Teaching experience in years	n	Mean	SD
0-4	12	6	6,4
5-14	38	11	8,3
15-24	13	5	6,6
25+	11	7	4,0
Total	74	9	7,6

## Conclusion and Discussion

This study revealed some interesting features about the practices of geography teachers regarding summative school-based examinations in pre-vocational education in the Netherlands and their relationship with meaningful learning. First, teachers rely largely on test items sourced from tests attached to the textbooks in constructing school-based examinations. Almost half the test items originate from these tests, and only approximately 30% of the test items come from other sources such as older national exams. According to the respondents, only 17% of the test items are self-constructed.

These results show that a rather small percentage of test items for internal school-based examinations devised by geography teachers in pre-vocational education in the Netherlands are self-constructed. From this study, it is unknown what the reason might be for this; it could be due to a lack of time, a lack of knowledge, or a low sense of self-efficacy regarding the construction of suitable test items. Although this is unknown for geography teachers in pre-vocational education in the Netherlands, results from research in other subjects in England suggest it might be a combination of lack of skills and lack of confidence (Black et al., 2010).

This is an important issue for further research because this study found some evidence that construction of test items by teachers does have a positive effect on their perception of test items' contributions to meaningful learning. Teachers who use more self-constructed test items perceive the percentage of test items contributing to meaningful learning as higher. Prior research showed that more than 60 percent of the test items in school-based examinations in pre-vocational education in the Netherlands are related to a type of remembering and, consequently, test items appealing to cognitive processes that transcend rote learning are less common in these examinations (Bijsterbosch, van der

Schee & Kuiper, 2016). Because teachers perceive the percentage of test items that contribute to meaningful learning as higher when those test items are self-constructed, self-construction of test items might be a promising principle changing teachers' practices with respect to classroom summative assessment.

In this respect, the fact that teachers with the longest teaching experience use more test items from tests attached to the textbooks is not a positive finding. The same relationship can be noticed with respect to the average age of geography teachers in pre-vocational education in the Netherlands. The older teachers become, the more they appear to rely on the tests attached to the textbooks.

Again, from this study it is unknown why older teachers and teachers with more teaching experience use more test items from tests attached to the textbook. Still, assuming that older teachers may function as role models for their younger colleagues, this is not a hopeful result. Functioning as a role model could be an important aspect for stimulating self-efficacy (Bandura, 1989; Schunk, 2003). When older teachers and teachers with more teaching experience rarely construct test items themselves, using them as models becomes problematic. These rather disappointing findings are even stronger when we consider that teachers who construct more test items themselves perceive the percentage of test items that contribute to meaningful learning as higher.

A note of caution with regard to these results is due here because it is unknown whether the teachers who use more self-constructed test items truly construct more test items that contribute to meaningful learning. The respondents were asked only to give an indication of the percentage of test items they thought were related to distinct cognitive processes. It might be the case that teachers who use more self-constructed test items think they use more test items that contribute to meaningful learning but in reality use the same percentage of test items that appeal to higher cognitive processes as teachers who perceive this percentage to be much lower.

Caution concerning the above results is also imperative because only 74 geography teachers in pre-vocational education responded to the questionnaire. Furthermore, the sample was not fully random. Although the invitation to fill in the questionnaire was published in multiple ways the chosen procedure might have affected the representativeness of the respondents.

Still, the construction of test items by teachers seems to have a positive effect on teachers' practices, beliefs and values regarding classroom summative assessments and their relationship with meaningful learning. This relationship with meaningful learning seems not to be affected, however, by other practices. No evidence was found for the effect to use instruments such as a test matrix or a taxonomy. Additionally, no evidence was found for a positive effect of working collaboratively, although we did expect some positive effects from such collaboration based on the literature (Harlen, 2005).

### **The Impact of high-stakes Tests**

In this study, some evidence was found for the impact of high-stakes tests on the internal school-based examinations. Almost three quarters of the teachers are convinced that they should use the same formats for test items in their internal school-based

examinations as are found in the external end-of-school (exit) examination because they feel that helps to prepare the students for the external end-of-school (exit) examination. Additionally, most teachers find it important to assess the objectives for the external end-of-school (exit) examination as well as the objectives for the internal school-based examinations. From these results the conclusion seems to be justified that a majority of the geography teachers in pre-vocational education are influenced by high-stakes tests such as the external end-of-school (exit) examination, not only with respect to the content validity of the internal school-based examinations but also in regard to formats for test items and the corresponding construct validity.

The results from this study suggest that teachers choose formats for test items that can be considered to give reliable test results. An approach to overcome these constraints between reliability and validity could be a dependability approach that emphasizes the reinforcing effect of both reliability and validity (Harlen, 2005). Dependability, in this sense, is the sum of reliability and validity and is meant to optimize reliability while ensuring validity, although this is not a calculable sum. This approach ensures the construct validity of the assessment while aiming at the highest possible reliability of the assessment scores (Harlen, 2004a; Wiliam, 1993). The application of test items and having well-specified criteria used to judge them is crucial when applying the concept of dependability (Harlen, 2005).

Black et al. (2011) confirm that this can be a helpful approach, especially when an appeal is made to the beliefs and values of teachers with respect to the purpose of summative classroom assessments. This approach of dependability, however, is not an easy one. Construct validity and reliability are often considered competing concepts, although classical test theory emphasizes that test validity can be reached only when the test scores are to some extent reliable (Berkel & Bax, 2006).

### **Teacher Professional Development**

The results from this study have given some input about how to accomplish change in geography teachers' practices, beliefs and values regarding the purpose and content of classroom summative assessment in pre-vocational education. To change teachers' practices, an enhancement of teachers' knowledge and skills on the relationship between assessment and meaningful learning seems to be needed. Furthermore, a change in the beliefs and attitudes of teachers regarding the purpose of summative classroom assessment seems necessary as well.

To stimulate change in teachers' practices and their beliefs, a professional development programme could be useful (Klenowski & Wyatt-Smith, 2011). Teacher professional development can be achieved in multiple ways. Guskey (1986, 2002) stressed the importance to start with teacher practices. When their practices can be changed, teachers will ultimately change their beliefs and attitudes as well.

By way of contrast, Clarke and Hollingsworth emphasized multiple pathways to achieve professional growth or development (Clarke & Hollingsworth, 2002). In their interconnected model of professional growth, it does not seem to be necessary to start with teachers' practices. Change in practices or in beliefs and attitudes can be fostered by

stimuli from outside and by reflection and enactment. Enactment of new classroom practices can stimulate change in teachers' beliefs but, on the other hand, a reflective change in beliefs and attitudes can result in new classroom practices as well.

To achieve the professional growth of geography teachers in pre-vocational schools in the Netherlands with respect to their practices, beliefs and values regarding classroom summative assessment further research is needed. Research into a teacher professional development programme in pre-vocational geography education in the Netherlands should provide insight into how to accomplish professional growth for teachers' practices, beliefs and values regarding classroom summative assessments and their relationship with meaningful learning.

## References

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental psychology*, 25(5), 729-735.
- Berkel, H., & Bax, A. (2006). *Toetsen in het hoger onderwijs* [Testing in higher education]. Houten: Bohn Stafleu van Loghum.
- Bijsterbosch, H., van der Schee, J. A., & Kuiper, W. (2016). Meaningful learning and summative assessment in geography education: An analysis in secondary education in the Netherlands. *International Research in Geographical and Environmental Education*. doi:10.1080/10382046.2016.1217076.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215-232.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451-469.
- Black, P., & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (1998b). *Inside the black box: raising standards through classroom assessment*. London: NferNelson.
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' Assessment Practices: Preparation, Isolation, and the Kitchen Sink. *Educational Assessment*, 3(2), 159-179.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and teacher education*, 18(8), 947-967.
- Guskey, T. R. (1986). Staff Development and the Process of Teacher Change. *Educational Researcher*, 15(5), 5-12.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: theory and practice*, 8(3), 381-391.

- Harlen, W. (2004a). A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W. (2004b). A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *Curriculum Journal*, 16(2), 207-223.
- Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1). In: *Research Evidence in Education Library*. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W., & Deakin Crick, R. (2003). A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- James, M., & Gipps, C. (1998). Broadening the basis of assessment to prevent the narrowing of learning. *Curriculum Journal*, 9(3), 285-297.
- Jo, I., & Bednarz, S. W. (2014). Dispositions Toward Teaching Spatial Thinking Through Geography: Conceptualization and an Exemplar Assessment. *Journal of Geography*, 113(5), 198-207.
- Klenowski, V., & Wyatt-Smith, C. (2011). The impact of high stakes testing: the Australian story. *Assessment in Education: Principles, Policy & Practice*, 19(1), 65-79.
- Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly*, 19(2), 159-172.
- Schussler, D. L. (2006). Defining Dispositions: Wading Through Murky Waters. *The Teacher Educator*, 41(4), 251-268.
- Schussler, D. L., Stooksberry, L. M., & Bercaw, L. A. (2010). Understanding teacher candidate dispositions: Reflecting to build self-awareness. *Journal of Teacher Education*, 61(4), 350-363.
- Stimpson, P. (2006). Changing assessments. In L. Lidstone & M. Williams (Eds.), *Geographical Education in a Changing World: Past Experience, Current Trends and Future Challenges*. (pp. 73-84). Dordrecht: Springer.
- Weeden, P. (2013). How do we link assessment to making progress in geography? In D. Lambert & M. Jones (Eds.), *Debates in Geography Education* (pp. 143-154). London: Routledge.
- Wertheim, J. A., Edelson, D. C., & The Road Map Project Assessment Committee (2013). A Road Map for Improving Geography Assessment. *The Geography Teacher*, 10(1), 15-21.
- William, D. (1993). Validity, dependability and reliability in National Curriculum assessment. *Curriculum Journal*, 4(3), 335-350.

### **Biographical Statements**

**Erik BIJSTERBOSCH** is geography educator at Windesheim University of applied sciences. His work focuses on curriculum, pedagogy and assessment in geography education.

**Joop VAN DER SCHEE** is professor for geography education appointed by the Royal Dutch Geography Society. He is working at VU University Amsterdam in the field of geography education. He is co-chair of the IGU Commission on Geographical Education. His research focuses on (digital) map skills and thinking through geography.

**Wilmad KUIPER** is professor for curriculum evaluation regarding science education appointed by the Netherlands Institute for Curriculum Development. He is working at Freudenthal Institute of Utrecht University in the field of curriculum development and curriculum evaluation. He is manager of the Research & Consultancy department of the Netherlands Institute for Curriculum Development.

**Tine BÉNEKER** is lecturer in geography education at Utrecht University and associate professor pedagogy of the school subjects geography, history, economics, social studies and religious studies at Fontys University of Applied Sciences. Her research focuses on the future dimension in education.