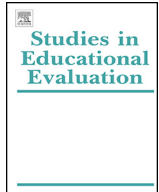




Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc



Formative use of test results: A user's perspective

Dorien Hopster-den Otter^{a,*}, Saskia Wools^b, Theo J.H.M. Eggen^a, Bernard P. Veldkamp^c

^a RCEC, Cito, University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands

^b Cito, Amsterdamseweg 13, 6814 CM, Arnhem, The Netherlands

^c RCEC, University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:

Received 29 June 2016

Received in revised form 10 October 2016

Accepted 17 November 2016

Available online xxx

Keywords:

Formative assessment

Test use

Test results

Information needs

Needs assessment

ABSTRACT

Despite the potential of using test data to support student learning, several studies have concluded that the actual use of test data remains limited. The present study addresses this problem by examining (1) the types of actions for which teachers, internal coaches, principals and parents within primary education want to use test results and (2) the information needed to perform these actions. The results obtained from the questionnaires show that the various users want to use test results for actions that support learning, which amounts to a discrepancy relating to actual use. Furthermore, the various users perform actions on different levels, thus indicating the need for tailored reports that fit the information needs of individual users. The results of the focus group method reveal the information needs of teachers, suggesting implications for the development of new score reports.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Research points to the potential of formative assessments as a way of supporting student learning (Baird, Hopfenbeck, Newton, Stobart, & Steen-Utheim, 2014; Black & Wiliam, 2009; Popham, 2009; Schildkamp & Lai, 2013). Formative assessments provide teachers with data about student performance. This data can be used to make decisions about the next steps in instruction, which are likely to be better, or better founded, than the decisions teachers would have taken intuitively in the absence of that data (Black & Wiliam, 2009).

To be able to use test data for student learning, teachers perform several cognitive steps (Davenport & Prusak, 1998; Ebbeler, Poortman, Schildkamp, & Pieters, 2016; Marsh, 2012). First, the collected data must be interpreted by giving meaning to scores. This can be done by summarizing the data in a more concise form. Subsequently, the interpreted data has to be contextualized by, for example, comparing the interpreted data with other information. The combination of different sources of information results in usable knowledge, which serves as a basis for decisions about an action, after which the action is executed. The impact of the action on student learning can then be evaluated using new data. As such, an iterative process is created (Mandinach & Jackson, 2012).

Several studies show that teachers have difficulty completing the phases of this iterative process (e.g. Hambleton & Slater, 1997; Hellrung & Hartig, 2013; Meijer, Ledoux, & Elshof, 2011; Schildkamp & Teddlie, 2008; Van der Kleij & Eggen, 2013). They especially struggle with (1) interpreting the test results and (2) translating them into actions that support learning. There are two possible explanations for these problems. First, the presentation regarding test results does not correspond with the assessment literacy skill level of teachers, resulting in difficulty interpreting the data and thereby making inappropriate use of the test results, with all its attendant consequences (e.g. Popham, 2009; Zapata-Rivera, VanWinkle, & Zwick, 2012). Second, the content of the presented data does not fit the information needs of teachers, resulting in problems translating the data into actions that support learning (e.g. Wiliam, 2011).

A considerable number of studies address the first explanation by allowing teachers and other users to develop the required assessment literacy skills (e.g. Lukin, Bandalos, Eckhout, & Mickelson, 2004; Verhaeghe, Vanhoof, Valcke, & Van Petegem, 2011). For example, some studies show a positive effect of training in terms of developing the required knowledge and skills to analyse and interpret data (e.g. Ebbeler et al., 2016; Van Geel, Keuning, Visscher, & Fox, 2016; Zwick et al., 2008). Other studies address the interpretation problem by adjusting the data presentation to the user's skill level (e.g. Van der Kleij, Eggen, & Engelen, 2014) since it has been suggested that the chosen method of data visualization can reduce the assessment literacy needs of users (Hattie & Brown, 2008).

* Corresponding author.

E-mail addresses: d.denotter@utwente.nl (D. Hopster-den Otter), saskia.wools@cito.nl (S. Wools), theo.eggen@cito.nl (T.J.H.M. Eggen), b.p.veldkamp@utwente.nl (B.P. Veldkamp).

<http://dx.doi.org/10.1016/j.stueduc.2016.11.002>

0191-491X/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The second explanation regarding the problem of using test data for student learning focuses on the content of the presented data. According to Zapata-Rivera and Katz (2014), everyone involved in the learning process of a child uses test results as presented through score reports, yet each audience has its unique types of decisions to be made on test results. If a score report designer defines the needs of the target audience, he opens up the possibility of tailoring the score report to meet the unique information needs of that audience. Within the target audience, four groups of users are distinguished: teachers, who are responsible for instruction and teaching a group of students; internal coaches, who coach teachers and support students with special needs across classes; principals, who are responsible for the school organization and parents, who support the learning of their own child.

Fitting the presented data with the information needs of users is often overlooked. According to Wiliam (2011), assessment data are made available to users under the assumption that this data are useful in some way. Too little attention has been paid to the types of actions that intended audiences want to perform on the basis of test data. The current study addresses this problem in the context of Dutch primary education. It seeks to determine the types of actions that teachers, internal coaches, principals and parents in primary education want to perform with the use of test results and the information needed to enable these actions.

1.1. Educational decision-making

In education, decision-making about instructional processes is an everyday activity. These decisions are taken at individual, group and school levels and can have important consequences for student learning. For example, on an individual level, decisions may pertain to whether a student should receive additional support. On a group level, decisions can relate to categorizing students into different levels for differentiation of instruction. On a school level, decisions may pertain to selecting a new teaching method. In order to ascertain whether these kinds of decisions are correct, it is important that decisions are informed by high-quality evidence (Brookhart & Nitko, 2008).

Test results are one source of data that can be used as evidence to support educational decision-making (Zapata-Rivera & Zwick, 2011). A test can be described as “an instrument or systematic procedure for observing and describing one or more characteristics of a student using either a numerical scale or a classification scheme” (Brookhart & Nitko, 2008; p. 5). Combined with other assessment data, such as student observations, oral questions and students’ work, an accurate picture of the student can be obtained and decisions can be informed (Brookhart & Nitko, 2008; Mandinach, 2012).

Despite the availability of test data meant to inform the didactical decisions of teachers, various studies conclude, however, that the actual use of test data for student learning is limited (Ledoux, Blok, Boogaard, & Krüger, 2009; Meijer et al., 2011; Vanhoof, Verhaeghe, Verhaeghe, Valcke, & Van Petegem, 2011; Verhaeghe et al., 2011). Instead, test data are used for other purposes, such as communication and evaluation, which do not automatically result in increased student learning. The use of data for communication has to do with informing parents about students’ ability or with informing inspectorate¹ for the purpose of accountability (Ebbeler et al., 2016; Van der Kleij & Eggen, 2013) while the sole purpose of the use of data for evaluation is to

appraise students’ performance. The actions that could follow from these judgments are not carried out (Brookhart & Nitko, 2008).

1.2. Presentation of test results

Test results are presented using score reports. Score reports are the vehicle for translating the test results into useful actions that support learning. It is a form of communication, with a sender, a message and an audience. The sender of score reports is the test developer or test agency presenting the results. The message deals with the content of the score report, and the audience consists of the people who use the test results (Hattie, 2009; Ryan, 2006).

To foster the use of test results for educational decision-making, the score report content should directly inform the audience about their decisions (Aschbacher & Herman, 1991; Hattie, 2009; Zapata-Rivera & Katz, 2014). Understanding the purpose for reading the test results in a score report helps to present the right message. Questions illustrating this statement include: What are the users’ goals? What do the users want to know? What decisions should the information inform, or what actions should it motivate or justify? If the score report presents content tailored to a user’s desired actions or decisions, the user would always know what to do with data that have collected and presented (Aschbacher & Herman, 1991; Wiliam, 2011).

1.3. Tailoring score reports to various users

Test results are often used by more than one intended audience, including teachers, parents, internal coaches and principals. As pointed out by Zapata-Rivera and Katz (2014) and Mandinach (2012), depending on the position of the user, each audience has its unique types of decisions to be made on the basis of test results. For example, teachers would be more involved in the decision process of an individual student or group of students while principals would be more focused on the decision process at the school level (Schildkamp & Kuiper, 2010). Internal coaches would be interested in the performance of all students while parents would be more interested in the performance of their own child (NEGP, 1998). With various intended audiences, it is likely that specially designed reports would be needed for each. The need for tailored reports will thus be reinforced depending on the variations among the decisions and information needs of the different audiences (Bradshaw & Wheeler, 2009; Hambleton & Slater, 1997).

1.4. Identifying users’ needs

It is the responsibility of test developers to ensure that the content of the score report fits the information needs of the user (Ryan, 2006). Because of this responsibility, various studies have called for the creation of score reports that meet the needs of different audiences (Aschbacher & Herman, 1991; Goodman & Hambleton, 2004; Hambleton & Slater, 1997; Jaeger, 1998; Wainer, Hambleton, & Meara, 1999). Hambleton and Zenisky (2013) and Zapata-Rivera et al. (2012) present a model for score report development – a user-centred model which starts with a needs assessment. This needs assessment should establish common ground between the test developer and the test user, bridging the gap between the information that results from an assessment and the actions the user wants to perform from the information. The results from the needs assessment will be the basis on “which all of the other steps in report design are linked” (Hambleton & Zenisky, 2013; p. 486).

The current study performed such a needs assessment. As mentioned earlier, its aim was to determine the types of actions that various users would like to perform with the use of test results as well as the information needed to enable these actions.

¹ The Dutch Inspectorate assesses and stimulates the quality of primary education and reports on the quality of each school to the public

Regarding this aim, we looked beyond the available information from existing tests and score reports. Instead, the focal point of this study was on starting from the decisions or actions that a user would ideally like to make (William, 2011) in order to support the use of test results for student learning.

1.5. Research questions

The main questions addressed in this study are as follows:

1. Which types of actions would users choose as desired uses of test results and how do these actions relate to actual uses?
2. What, if any, is the extent of the differences between teachers, internal coaches, principals and parents with regard to desired and actual uses and corresponding actions?
3. What information from test results is needed to perform the desired actions?

2. Method

In order to answer the research questions, data were collected from different user groups within Dutch primary schools. Four different kinds of users were distinguished: teachers, internal coaches, principals and parents and guardians (hereafter, parents). A questionnaire was developed for the teachers, internal coaches and principals to identify the actions for which test results are used in the context of teaching. As the focus was on actions related to teaching and were, therefore, not applicable to parents, a separate questionnaire for parents was developed.

In addition, qualitative data from focus groups were gathered to validate the results of the questionnaire data and to further specify the information needs. Based on the results from the questionnaires, we decided to target these focus groups at teachers. In order to facilitate the readability of this article, these choices are further elaborated in the results section.

Table 1 shows the relation between the research instruments and the different user groups and research questions. In the next session, the instruments, procedure, data analyses and sample characteristics are discussed.

2.1. Instruments and procedure

2.1.1. Questionnaire 1–teachers, internal coaches and principals

The first questionnaire was developed to investigate the actions that teachers, internal coaches and principals deemed desirable in relation to test results within the context of teaching as well as the actual use of such results. Test results were defined as results from a systematic instrument, such as written or digital tests, excluding results from other assessment methods like observations and verbal responses from students. The actual use depends on the availability of information from current tests, which in the Netherlands, are mostly standardized tests aimed at monitoring students and written or digital tests from teaching methods. In terms of desired use, respondents were asked to mention all actions independent of currently available information and tests.

Alongside the eleven items on the background of the respondents, the questionnaire consisted of three items showing a list of possible actions for which test results could be used. The first and second items consisted of multiple response questions in which respondents were required to select actions relating to actual and desired use, respectively. As respondents could select all actions as desired use, the third item asked respondents to choose the most important desired action from their selection. This provided greater insight into the degree of interest relating to the different actions. The questionnaire is included in Table A.1 of Appendix A.

The list of possible actions resulted from the grid shown in Table 2. This grid consists of actions on three levels (individual level, group level and school level) and three purposes (communicating learning, supporting learning and evaluating learning). This enabled the possibility to describe some patterns in the answers. The levels were related to the precise data used for the action. For example, the placement of students into groups for differentiation is a group level action because the data from the student group is used to perform this action. Purposes refer to what data is used for. For example, determining individual students’ performance compared to the national performance is meant to appraise student performance without setting new learning goals. Therefore, this action is labelled as having an evaluative purpose. Although this study is primarily aimed at actions that support learning, the purposes of communicating and evaluating learning were added in order to gain a better understanding of total actual use.

The grid containing possible actions is based on the questionnaire of Blok, Otter and Roeleveld (2001). They collected a list of actions for which data from various tests could be used. We validated the actions from the grid (Table 2) by asking two educational consultants, specialized in assessments, to generate as many actions as they could think of for which teachers, internal coaches and principals would like to use test results. These two educational consultants have considerable contact with all the various user groups about their desired and actual uses of tests. The mentioned actions were already included in the questionnaire. We also asked three teachers to describe what is meant by each action or to give an example of an action from their own practice. We concluded that the descriptions of the actions were clear and appropriate to the Dutch context. We pretested the questionnaire by asking five teachers to fill out the questionnaire and to indicate whether they were missing some actions or whether some questions were unclear. This resulted in a few minor adaptations such as the addition of the option “no use of test results”. This option ensured a distinction between respondents who did not answer the question because they skipped it and those who did not make use of test results. Furthermore, we added the option “other” so that the respondents could mention actions outside the list.

The teachers, internal coaches and principals were asked to complete the questionnaire through various channels. Those schools opting to participate as a focus group also received an e-mail with a link to the questionnaire, which was distributed within the school. The questionnaire was filled out electronically by the respondents.

Table 1
Relation between instruments, respondents and research questions.

Instruments	Respondents	Research questions		
		1 actions	2 user differences	3 information needs
Questionnaire 1	Teachers, internal coaches, principals	x	x	
Questionnaire 2	Parents	x	x	
Focus group	Teachers	x		x

Table 2

Grid consisting of three levels and three purposes of possible actions for which test results can be used.

Levels	Purposes		
	Communicating learning	Supporting learning	Evaluating learning
Individual level	<ul style="list-style-type: none"> - To inform parents/guardians during individual meetings or by means of score reports - To inform the individual student about his/her performance - To inform other schools about an individual student by means of an educational report (school transition of the student) 	<ul style="list-style-type: none"> - To create individual action plans for low performing students - To create individual action plans for high performing students - To give feedback to students in order to formulate their own learning goals 	<ul style="list-style-type: none"> - To determine individual students' performance compared to the national performance - To determine individual students' progress regarding learning goals or content - To make decisions about students' transition year
Group level	<ul style="list-style-type: none"> - To inform parents/guardians during group meetings - To inform the student group about their performance - To inform colleagues about the student group during a group discussion or transmission 	<ul style="list-style-type: none"> - To create group action plans - To adapt instruction to educational needs - To place students into different groups for differentiation 	<ul style="list-style-type: none"> - To determine the group performance compared to the national performance - To determine group progress regarding learning goals or content - To compare parallel groups regarding their progress
School level	<ul style="list-style-type: none"> - To inform people via the school prospectus or school website - To inform the school board or participation council - To inform the inspectorate 	<ul style="list-style-type: none"> - To create school or annual plans - To formulate policy regarding the selection of new teaching methods - To create professional plans (performance appraisals, career decisions) 	<ul style="list-style-type: none"> - To determine the school's performance compared to the national performance - To determine progress regarding school goals - To compare the performance of a student group with (those of) previous years.

2.1.2. Questionnaire 2–parents

The questionnaire used for the parents addressed the perspective of parents on the use of test results. Alongside six items about parents' background, the questionnaire consisted of four items. This article addresses only the two items relating to the purpose of using tests and actions aimed at supporting the learning process. The first item was a multiple choice question in which parents had to select the purpose (communicating, supporting or evaluating learning) that best suited the reason for which they thought test results would be mostly used. The second item was a question investigating the supporting actions parents would like to take in order to determine the extent to which these actions differed from the actions of other users. The questionnaire is included in Table A.2 of Appendix A.

We pretested the questionnaire by asking three parents to fill out the questionnaire and to indicate whether some questions were unclear. The parents reported that all questions were clear, which was also demonstrated in the responses they provided. Therefore, the pretest did not result in adaptations to the questionnaire.

Parents were asked to complete the questionnaire using various channels. Some schools agreed to participate as a focus group and distributed the questionnaire to the parents of their school. Another example was a call on an educational website for parents. The questionnaire was filled out electronically.

2.1.3. Focus group–teachers

Focus group meetings with the teachers of seven participating schools were held to validate the results of research question 1 and to identify the information needs for question 3. The design of the focus group method included the characteristics of a group interview as well as a group discussion (Newby, 2010).

The meeting consisted of three parts. In the first part, the results from the questionnaire were validated by identifying the purpose of teaching and the corresponding actions aimed at achieving this purpose. While the questionnaire was a reactive task whereby respondents were asked to select actions from the given list, the focus group was a generating task whereby respondents were asked to identify the actions by themselves. In the second part, the researcher discussed some conceptual aspects of formative

assessments in order to achieve the same understanding of the concept. In the third part, the teachers were to select the actions from the first part for which they needed information from formative assessments. Actions that did not require information were not selected. To illustrate, the teachers selected the action "placement of students into different instruction groups for differentiation" and not the action "using humor" because they needed some information for the first action but not for the second. Thereafter, the teachers had to think about the information needed for each action. All individual answers were recorded on paper. The teachers' responses were then systematically grouped and validated during the focus group.

The structure of the focus group meeting and the formulation of the questions were first pretested using individual interviews with two respondents who did not participate in the focus groups. These interviews resulted in some adaptations regarding the formulation of the questions; for example, we changed the following question "which information *from tests* do you need in order to carry out this action?" into "which information do you need in order to carry out this action?" This is because the pretest showed that respondents only gave answers about information that known test reports are able to give, a mindset deemed too limited for this study. In addition, the first focus group was meant as a trial. Since no changes were made afterwards, the data from this focus group were included.

2.2. Data analysis

2.2.1. Questionnaire 1–teachers, internal coaches and principals

To answer the first question, frequency analyses were used to show the number of occurrences of each response chosen by the respondents. Because the questionnaire included two multiple response questions, the number of responses differed from the number of respondents. We used McNemar's test to ascertain statistical differences between actions in terms of actual and desired use. This test evaluates the difference between two correlated proportions, which means that the two scores are not independent. In order to describe the patterns in the answers, the number of actions was then summarized into the three purposes: communicating, supporting and evaluating learning.

These “purpose” subscales all had high reliabilities relating to desired use, with Cronbrach’s α of 0.80, 0.81 and 0.86, respectively. Regarding actual use, the reliabilities of these subscales were moderate, with Cronbrach’s α of 0.59, 0.65 and 0.74, respectively. For the second question, the same analyses were performed, but we divided the respondent group into teachers, internal coaches and parents. In addition, the number and percentages of actions were summarized into three levels: individual, group and school. Since there were no additional actions mentioned under the option “other” from outside the given list, we did not analyse these answers.

2.2.2. Questionnaire 2–parents

The data from the parent questionnaire were analysed both qualitatively and quantitatively. The quantitative analysis consisted of frequency analyses of those questions with a closed-answer format. The answers to the open-answer question were coded in a qualitative way. We compared the answers of the open-answer questions and then grouped related pieces of information into categories. We subsequently used these categories to classify all answers. If an answer did not fit into the existing categories, the framework was modified and the process repeated.

2.2.3. Focus group–teachers

The participants’ responses to the questions in the focus group meetings were listed, grouped and documented during the focus group meeting. For the analyses, answers were considered irrelevant and were removed if they did not correspond with actions and needed information. For example, some teachers mentioned a method of testing (e.g. doing an observation) or some preconditions regarding teaching their students (to create an orderly group climate). Following the focus group method, all relevant data were regarded as valuable, regardless of how many teachers appointed the data. The results of the different focus groups were summarized and compared.

2.3. Sample characteristics

2.3.1. Questionnaire 1–teachers, internal coaches and principals

A total of 140 teachers, 34 internal coaches and 14 school principals filled out the questionnaire. Of these 188 respondents, 30 respondents did not complete the questionnaire, which means that the responses of 158 respondents were used for analysis. Some background characteristics of the 158 respondents are presented in Table 3. The sample characteristics are typical of the Dutch primary school teacher population (www.onderwijscijfers.nl).

2.3.2. Questionnaire 2–parents

Altogether, 250 parents of students in primary education participated in this study. However, 33 parents did not complete the questionnaire, which means that the responses of 217 parents (48 males, 169 females) were used for analysis. The distribution

relating to the educational level of the respondents was 60% completing higher education, 29% completing vocational education, and 11% had obtained a lower educational level. Overall, the sample included a relatively high proportion of female and highly educated respondents compared to the population of parents in the Netherlands.

2.3.3. Focus group–teachers

Focus groups were held at seven different schools. All teachers within a school participated in the corresponding focus group. We could therefore ensure that the data were gathered from enthusiastic teachers as well as those who were not very enthusiastic about using tests. To further enhance the representativeness, we selected schools of different sizes. The school teams varied between seven and 17 persons. In total, 84 teachers participated in the seven focus groups. We have no reason to believe that this sample does not reflect the characteristics of the school population.

3. Results

3.1. Question 1: which types of actions would users choose as desired uses of test results and how do these actions relate to actual uses?

The 158 respondents indicated 1922 actions as desired uses of test results (Table 4), representing, on average, more than 12 actions per respondent. The most frequently chosen action under desired use was “to inform parents during individual meetings or by means of score reports” selected by 121 respondents (76.6%). This action accounted for 6.3% of all the desired use answers. Informing parents was also the most frequently chosen action under actual use (91.1%). However, this action was selected significantly less frequently as a desired action than as an actual use ($\chi^2 = 14.7$; $p < 0.001$). Communications to the inspectorate, the creation of group plans and some actions relating to the evaluation of test results were also selected significantly less frequently as desired use than as actual use. The creation of group plans was still the second most frequently chosen desired use action (72.2%).

Some actions were chosen significantly more often as desired uses than as actual uses. For example, the frequency of the action “to give feedback to students in order to formulate their own learning goals” doubled ($\chi^2 = 46.2$; $p < 0.001$) from 19.6% to 51.3%. Other examples included the creation of action plans for high performing students ($\chi^2 = 5.8$; $p = 0.02$) and the formulation of policy regarding the purchase of teaching methods and instruments ($\chi^2 = 19.2$; $p < 0.01$).

Other frequently mentioned actions as desired uses, although they were not chosen significantly more frequently, were the creation of individual action plans for low performing students (70.3%) and the placement of students into groups for differentiation (68.4%). These actions, including the creation of group plans (72.2%), were all examples of actions relating to the category of supporting learning.

Table 5 presents a summary of the number and percentages of actions into the three purposes: communicating, supporting and evaluating learning. The action “no use of test results” was a separate category that did not belong under any of the other three purposes. Notwithstanding the fact that fewer actions were selected as desired use ($n = 1922$) in comparison with actual use ($n = 1991$), the number of actions relating to supporting learning was higher for desired use ($n = 741$) than for actual use ($n = 656$). The opposite was true for the purposes of communicating and evaluating learning. Regarding the relative distribution of desired use, respondents mostly chose actions relating to supporting learning (38.6%). This result differed from actual uses whereby actions relating to the purpose of evaluating learning were most

Table 3
Background characteristics of respondents (N = 158) from questionnaire 1.

	Teachers	Internal coaches	Principals	Total
Sex				
Male	19	1	6	26
Female	100	26	6	132
Average age (SD)	40.4 (11.3)	45.9 (8.8)	49.8 (10.4)	49.8 (10.4)
Years of total experience				
0–5	17	3	0	20
5–10	26	1	0	27
10<	76	23	12	111
Total	119	27	12	158

Table 4
Number of responses and respondents (N = 158) choosing an action as desired use and actual use.

Actions	Desired use			Actual use		
	Responses		Respondents	Responses		Respondents
	n	%	%	n	%	%
Communicating learning						
* To inform parents/guardians during individual meetings or by means of score reports	121	6.3	76.6	144	7.2	91.1
* To inform the individual student about his/her performance	83	4.3	52.5	68	3.4	43.0
To inform other schools about an individual student by means of an educational report (school transition of the student)	89	4.6	56.3	102	5.1	64.6
To inform parents/guardians during group meetings	29	1.5	18.4	27	1.4	17.1
* To inform the student group about their performance	39	2.0	24.7	22	1.1	13.9
To inform colleagues about the student group during a group discussion or transmission	100	5.2	63.3	113	5.7	71.5
To inform people via the school prospectus or school website	15	0.8	9.5	13	0.7	8.2
To inform the school board or participation council	32	1.7	20.3	42	2.1	26.6
* To inform the inspectorate	46	2.4	29.1	88	4.4	55.7
Supporting learning						
To create individual action plans for low performing students	111	5.8	70.3	115	5.8	72.8
* To create individual action plans for high performing students	104	5.4	65.8	86	4.3	54.4
* To give feedback to students in order to formulate their own learning goals	81	4.2	51.3	31	1.6	19.6
* To create group action plans	114	5.9	72.2	137	6.9	86.7
To adapt instruction to educational needs	101	5.3	63.9	91	4.6	57.6
To place students into different groups for differentiation	108	5.6	68.4	118	5.9	74.7
To create school or annual plans	39	2.0	24.7	32	1.6	20.3
* To formulate policy regarding the selection of a new teaching method	63	3.3	39.9	32	1.6	20.3
To create professional plans (performance appraisals, career decisions)	20	1.0	12.7	14	0.7	8.9
Evaluating learning						
* To determine the individual students' performance compared to the national performance	67	3.5	42.4	97	4.9	61.4
To determine individual students' progress regarding learning goals or content	107	5.6	67.7	118	5.9	74.7
* To make decisions about students' transition year	64	3.3	40.5	82	4.1	51.9
* To determine the group performance compared to the national performance	71	3.7	44.9	89	4.5	56.3
To determine group progress regarding learning goals or content	92	4.8	58.2	90	4.5	57.0
To compare parallel groups regarding their progress	33	1.7	20.9	28	1.4	17.7
To determine the school's performance compared to the national performance	58	3.0	36.7	55	2.8	34.8
To determine progress regarding school goals	62	3.2	39.2	66	3.3	41.8
* To compare the performance of a student group with (those of) previous years.	65	3.4	41.1	89	4.5	56.3
No use of test results	8	0.4	5.1	2	1.0	1.3
Total	1922	100		1991	100	

Note: *p < 0.05.

commonly chosen (35.9%), and actions relating to the purpose of supporting learning were chosen less frequently (32.9%).

The results shown in Table 5 were confirmed by the answers on the third questionnaire item, which required respondents to choose the most important action as a desired use of tests. In total, 53.9% of the respondents chose an action relating to the purpose of supporting learning. The most frequently chosen action in this category was "to adapt instruction to educational needs" (n = 25), followed by "to give feedback to students in order to formulate their own learning goals" (n = 18) and "to create group action plans" (n = 16). Actions relating to the purpose of evaluating learning were chosen by 27.5% of the respondents. This result was mainly due to the action "to determine individual students' progress regarding learning goals or content" (n = 29, 17.4%).

The view of parents corresponded with this result; 45.2% of them indicated that test results were mainly used to support their

Table 5
Number and percentage of actions relating to desired and actual use per purpose.

Purpose	Desired use		Actual use	
	Count	%	Count	%
Communicating learning	554	28.8	619	31.1
Supporting learning	741	38.6	656	32.9
Evaluating learning	619	32.2	714	35.9
No use of test results	8	0.4	2	0.1
Total	1922	100	1991	100

child's learning. This was followed by 40.1% of parents, who thought that student-level evaluation was the main goal, and 14.7% who said that communicating results to the parents, principal or inspectorate was the central goal.

Based on these results, we conclude that respondents mostly chose actions relating to the purpose of supporting learning, which amounts to a discrepancy relating to actual use. In order to create useful score reports of test results, we investigated whether these actions were the same or different for the various audiences.

3.2. Question 2: what, if any, is the extent of the differences between teachers, internal coaches, principals and parents with regard to desired and actual uses and corresponding actions?

Table B.1 in Appendix B presents the percentages of teachers, internal coaches and principals choosing an action for desired and actual use. No major differences were found with regard to the three purposes of communicating, supporting and evaluating learning (Table 6). Teachers and principals mostly chose actions as desired use under the purpose of supporting learning while internal coaches chose almost as many actions for the purpose of supporting learning as for the purpose of evaluating learning. All user groups indicated that current test results were primarily used to evaluate learning.

There were, however, differences between the user groups with regard to the different levels of actions (Table 7). The teachers especially selected actions relating to the individual level (45.5%) and subsequently chose many actions relating to the group level

Table 6
Response percentages of actions chosen by users in relation to the different purposes.

Purpose	Desired use			Actual use		
	Teachers (n = 119)	Internal coaches (n = 27)	Principals (n = 12)	Teachers (n = 119)	Internal coaches (n = 27)	Principals (n = 12)
Communicating learning	28.6	28.5	31.5	31.5	28.6	33.7
Supporting learning	39.8	35.3	37.6	33.8	31.0	31.1
Evaluating learning	31.1	36.0	30.4	34.6	40.4	35.2
No use of test results	0.5	0.2	0.6	0.1	0.0	0.0
Total	100	100	100	100	100	100

Table 7
Response percentages of actions chosen by users in relation to the different levels.

Level	Desired use			Actual use		
	Teachers (n = 119)	Internal coaches (n = 27)	Principals (n = 12)	Teachers (n = 119)	Internal coaches (n = 27)	Principals (n = 12)
Individual	45.5	39.9	33.1	44.5	40.7	31.2
Group	37.1	34.0	30.4	38.1	32.2	28.1
School	16.9	25.9	35.9	17.3	27.1	40.7
No use of test results	0.5	0.2	0.6	0.1	0	0
Total	100	100	100	100	100	100

(37.1%). Only a small number of teachers' responses represented actions relating to the school level (16.9%). The answers furnished by the internal coaches showed a similar pattern although they had a greater preference than teachers to perform some actions at the school level (25.9%). The principals' answers showed the opposite, with most selected actions relating to the school level (35.9%). To illustrate this difference, at the school level, the development of school plans was selected far more frequently by principals (91.7%) than by teachers (13.4%) and internal coaches (44.4%).

Parents (N = 217) mentioned also actions relating to supporting the learning process. Helping their child with homework was, for example, the most mentioned action (19.0%). Furthermore, 17.2% of the parents would like to practice the learning material with their child at home. Some parents would give their child additional support by providing learning material to remedy weaknesses (14.5%). Other examples of actions mentioned included reading books (9.3%), testing their child on his/her knowledge for a test (7.9%), learning in a playful way (4.5%), helping to develop learning skills like planning school work (4.1%), giving some educational games (3.4%) and visiting cultural organizations like museums (3.4%). All such actions were in relation to their own individual child.

These differences in actions between the various users indicated that there is a need for score reports to be tailored to the specific user groups, corresponding to the actions that these kinds of users would like to undertake (Zapata-Rivera & Katz, 2014). This means that we should investigate the information needs of each user group separately. Based on the previous finding that test results would rather be used to support learning, we decided to limit our focus on teachers for the third question. Teachers' primary task was to support the learning process of students. They were also the users who actually communicated these results to other users such as students and their parents.

3.3. Question 3: what information from test results is needed to perform the desired actions?

The results of research question 1 were validated during the seven focus group meetings. The teachers in the focus groups underlined the general principle that they would support student learning by developing the cognitive and social knowledge and skills of their students.

Subsequently, the actions for achieving this purpose were generated. Nine actions in the questionnaire were related to the

purpose of supporting learning. The most frequently chosen actions in the questionnaire were also generated by the teachers in the focus groups. Starting with the most frequently mentioned action, these actions were: (1) alignment of learning material and learning objectives with the initial level of students, (2) placement of students into different instruction groups for differentiation, (3) student-teacher conversations about well-being and learning, (4) development of group and individual action plans and (5) alignment of learning materials to learning objectives and preferences, with action (4) from the focus group covering three actions from the questionnaire. Two actions from the questionnaire were not mentioned in the focus group. However, these actions were also chosen less frequently by teachers but more often by principals and internal coaches. As the actions were formulated in the focus groups, the conceptualizations of these actions were slightly different from the description in the questionnaire. For example, student-teacher conversations about well-being and learning was related to the questionnaire action "to give feedback to students in order to formulate their own learning goals".

Table 8 presents the five actions and the corresponding information needs of each action. The action mentioned by all seven focus groups was the alignment of learning materials and learning objectives with the initial level of students, which corresponds with the questionnaire item "to adapt instruction to educational needs". In order to perform this action, teachers need information about the learning objectives for each year and subject as well as information about students' mastery of these learning objectives. Furthermore, information is needed with regard to the sequence of acquiring learning objectives, realistic expectations for the next learning objective and learning material suggestions of how to achieve this objective.

Most of the information needs mentioned for student-teacher conversations were about well-being and learning. For this action, teachers need information about students' learning in order to give students feedback, for example, information about students' strategy to solve assignments. Furthermore, teachers need information about students' personal aspects, like students' well-being, working attitude and self-efficacy.

The overall results presented in Table 8 show that teachers have different information requirements for performing actions: on one hand, information about general teaching aspects like the learning objectives for each year and subject and, on the other hand, information about students such as their learning

Table 8
Information needs mentioned by focus groups (N = 7) to perform actions to support learning.

Information needs	Actions				
	1	2	3	4	5
Learning material suggestions				■	■
Learning objectives for each year and subject	■				
Realistic expectations for next learning objective	■		■	■	
Sequence of acquiring learning objectives	■		■		
Starting level of students' knowledge					■
Students' interest in subjects					■
Students' learning preferences					■
Students' learning progress				■	■
Students' mastery of a learning objective	■	■	■	■	■
Students' motivation			■	■	
Students' self-efficacy			■	■	
Students' strategy to solve assignments			■	■	
Students' strong and weak points			■	■	
Students' well-being			■	■	
Students' working attitude			■	■	
Suggestions for placement of students into three groups		■			■

1=Alignment of learning material/objectives to initial level – questionnaire item: to adapt instruction to educational needs (n=7); 2=Differentiation – questionnaire item: to place students into different groups for differentiation (n=6); 3. Student-teacher conversations – questionnaires item: to give feedback to students in order to formulate their own learning goals (n=5); 4=Group and individual action plans – questionnaire item: to create group/individual action plans for low/high performing students (n=5); 5=Alignment of materials to learning objectives/preferences – questionnaire item: to formulate policy regarding the selection of new teaching methods (n=3).

progress. Furthermore, teachers need information about both the personal aspects of students, like their interest in subjects, and their cognitive aspects, such as their mastery of a learning objective.

These results also indicate that teachers need the same information for different actions. For example, students' mastery of a learning objective was needed to perform four actions. Realistic expectations regarding subsequent learning objectives were mentioned for three actions. Other kinds of information were only selected for one action, like students' working attitude.

Because we started the focus group by discussing the actions for the purpose of teaching rather than the more specific actions relating to test results, and because we asked the focus groups for actions requiring information from all possible sources, including tests, teachers also mentioned three actions that were not directly related to the actions listed in the questionnaire: (1) placement of students into different groups for cooperative learning and collaboration, (2) connection to students' perception of the world and (3) creation of ownership. Cooperative learning and collaboration means that a group of students have to cooperate equally on assignments in order to achieve learning goals. For this action, teachers need information about the learning objectives for each year and subject and about students' mastery of these learning objectives. This is the same information as that mentioned earlier. However, teachers also need information about students' behavior, their social and communicative skills and their willingness to collaborate. The second action concerned the connection to students' perception of the world, which contained the adoption of the chosen examples, thematic topics and the introductions of lessons relating to students' experiences and interest. For this action, teachers also need additional information, such as the dynamic of a student group, the proportion of boys and girls and students' home situation. The third action concerned giving students responsibility to support their own learning. The information requirements included students' persistence and ability to work independently.

4. Conclusions and discussion

This study investigated the types of actions users want to perform with the use of test results and the information needed to enable these actions. By administering two questionnaires and conducting seven focus group meetings, both qualitative and quantitative data were gathered. In the analyses, distinctions were made among various users, including teachers, internal coaches, principals and parents.

The results of this study suggest that in relation to desired uses, respondents mostly chose actions relating to the purpose of supporting learning. The study also showed that this desired use of test results was not the same as the actual use; test results were primarily used to evaluate the learning process by determining the student's ability. These results corroborate the results of previous studies, suggesting the limited use of test results for formative purposes (Ledoux et al., 2009; Meijer et al., 2011; Vanhoof et al., 2011; Verhaeghe et al., 2011).

Furthermore, we conclude that the various users want to perform actions on different levels and in different contexts. Teachers and parents reported that they want to perform actions at the level of the individual student whereby teachers act in an educational setting and parents perform in a more informal situation. Internal coaches and principals selected more actions relating to the school level. This result is in accordance with the expectation regarding the unique decisions of each user group (Zapata-Rivera & Katz, 2014).

Based on the results of the first and second questions, we decided to limit our third question to teachers. The results from the first question were validated, and we gathered insights about the information needs of teachers to perform each action mentioned. The results show the need for different kinds of information, for instance, relating to students' strategy to solve an assignment, students' motivation and their working attitude. This result confirms Brookhart and Nitko's (2008) and Mandinach's (2012) argument that test data are only one source of information in supporting educational decision-making and that an accurate picture from the student could be obtained with the use of other assessment data. The results also indicate that teachers sometimes need the same information for different actions; for example, information about students' mastery of a learning objective was mentioned for the performance of four actions.

Finally, the formulation of the question "what information do you need in order to carry out this action?" expanded the mindset of respondents but also resulted in information needs which might not arise from tests. For example, the information need "sequence of acquiring learning objectives" likely formed a greater part of the content knowledge of the teacher. This illustrates the view of Gummer and Mandinach (2015) that the process of using test data is complex and that for instructional decision-making, teachers need to combine an understanding of data with "standards, disciplinary knowledge and practices, curricular knowledge, pedagogical content knowledge, and an understanding of how children learn" (p. 2).

4.1. Limitations of the study

This study was limited in several ways. First, the sample size was limited (especially the number of principals and internal coaches), so the results can only be generalized to a limited degree. However, the fact that most of the focus groups mentioned the same type of actions, which were also in the questionnaire, suggests that we have identified the most important actions for teachers.

The three additional actions mentioned in the focus groups suggest that the actions list in the questionnaire might have been incomplete. This is because we asked for actions from two different perspectives. We started the focus groups by discussing the actions for the purpose of teaching rather than the more specific information needs relating to test results. Furthermore, we asked the focus groups for actions requiring information from all possible sources, including tests. Thus, the answers from the focus group included actions from a wider perspective. Moreover, no additional actions were mentioned during the pretesting of the questionnaire; the “other” option in the questionnaire was not used; and two of the additional actions were mentioned by just two focus groups. For this reason, we considered this difference of actions to be of minimal importance.

Finally, the users chose the actions for which they want to use test results, but this choice was made within an existing frame of reference consistent with existing tests in the current national system of the Netherlands. This imagination seems to be difficult and contextualized, which limits generalization.

4.2. Implications for practice

This study provided insights into the actions and corresponding information needs of teachers. The results show that teachers and others would like to use test results for uses for which current measurement instruments are not validated. This may result in misuse or limited use of current test results. The results are informative to teachers and others, in terms of the use of different instruments, for their educational decisions about actions as the information needs of teachers cannot be obtained from one test. Furthermore, test developers could use the insights herein for the development of tests and score reports aimed at teachers. If the score report presents content that is tailored to the actions that teachers would like to undertake, then teachers would always be able to perform the action once the data is collected and presented. For example, teachers can use data from test results to make up different groups for differentiation, to align the learning objective with the initial student level or to develop group action plans. Compared to the available information from current tests, test developers should develop tests that offer more detailed information like student strategies to solve assignments. In this way, test results will be used to make decisions about the

next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken intuitively in the absence of such data (Black & Wiliam, 2009). The development of tailored score reports can contribute to the potential of formative assessment as a way of supporting student learning.

4.3. Implications for future research

The results and limitations imply a future research agenda. First, it seems worthwhile to examine the actions and information needs of students as a user group. Since teachers indicated that they would formulate learning objectives together with students more often than they actually did, it would be useful to also examine the actions and information needs of students. This result is in accordance with the trend towards activating students as owners of their own learning as a key strategy of formative assessment (Wiliam & Thompson, 2007).

Second, despite the growing body of research on effective score reporting (Zenisky & Hambleton, 2012), there has been little effort on users’ actual use of developed score reports. Future research is needed on how to design effective score reports for teachers that visualize test results that are appropriate to the identified actions and information needs reported in this study. Moreover, it would be useful to study the extent to which the presentation of the identified information needs result in more data use for student learning.

Third, the results showed that teachers need detailed information from tests, such as the extent to which each student has mastered a certain learning objective. This implies that score reports should visualize smaller levels of information (e.g. from total test scores to subscores and items). Accuracy is however related to the level of reporting. When reports are more detailed, the accuracy of test scores is often negatively impacted; e.g. accuracy is lower, and scores are more uncertain (Monaghan, 2006; Ryan, 2006). The total score is often a more accurate measure of an individual’s knowledge or skills in a subdomain of interest than a subscore derived only from those items that purport to measure the subdomain directly (Monaghan, 2006). Assessment organizations have a duty to provide teachers with sufficient information about these accuracies to allow them to make valid inferences based on test results (e.g. AERA et al., 2014; Newton, 2005). Future research is needed to investigate how to best communicate this accuracy information.

Appendix A.

Table A.1
Questionnaire 1.

Respondents’ background

1. What is your position within the school?
a) Teacher; b) Internal coach; c) Principal
2. What is your sex?
a) Male; b) Female
1. What is your age?
4. In which district do you work?
a) Friesland; b) Groningen; c) Drenthe; d) Overijssel; e) Gelderland; f) Limburg; g) Noord-Brabant; h) Zeeland; i) Utrecht; j) Noord-Holland; k) Zuid-Holland; l) Flevoland
5. How many years’ experience do you have in primary education? (For internal coaches/principals only)
a) Less than 5 years; b) 5 to 10 years; c) More than 10 years
6. How many years have you worked as a teacher/internal coach/principal?
a) Less than 5 years; b) 5 to 10 years; c) More than 10 years
7. Which grade do you teach the most? (Note: Answer the remaining questions for the students in this grade) (For teachers only)
a) Lowest grades (Group 1,2); b) Middle grades (Group 3,4,5); c) Upper grades (Group 6,7,8)
8. In which grade do you function as internal coach? (For internal coaches only)
a) Lowest grades (Group 1,2); b) Middle grades (Group 3,4,5); c) Highest grades (Group 6,7,8); d) Whole school (Group 1–8). Other:
9. What, if any, other functions do you fulfill in school? (For internal coaches/principals only)
a) No other functions; b) Principal; c) Teacher; d) ICT coordinator; e) Language or math specialist; f) Otherwise, namely . . .
10. Does your school use any of the following principles?
a) Anthroposophy; b) Dalton; c) Freinet; d) Jenaplan; e) Montessori; f) O4NT; g) none of these
11. Which of the following student monitoring systems do you use?
a) Cito-LVS; b) Parnassys; c) Esis; d) Dot.com; e) Other:

Table A.1 (Continued)

Actual use of test results

12. Test results may be used for different actions. We presented a number of actions below. Check the purposes for which you have used test results in recent school years (2014–2015, 2015–2016). Note: we mean the use of test results for actions with the majority of students, not for exceptional circumstances. It is possible to give more than one answer. In the last school years, I have used test results for the following actions . . .

Desired use of test results

13. Suppose you were allowed to design test score reports yourself so that it presents you with all the information you need, check for which actions you would like to use test results. Note: we mean the use of test results for actions with the majority of students, not for exceptional circumstances. It is possible to give more than one answer. In an ideal situation, I would like to use test results for the following actions . . .

Most important purpose of desired use

14. We presented your chosen actions regarding test results below. Which action do you find most important in the use of test results?

Chose the most important one.

For questions 9, 10 and 11, we presented the following action list:

- | | |
|---|--|
| No use of test results | To adapt instruction to educational needs |
| To inform parents/guardians during individual meetings or by means of score reports | To determine group progress regarding learning goals or content |
| To create individual action plans for low performing students | To inform the school board or participation council |
| To determine individual students' performance compared to the national performance | To formulate policy regarding the selection of new teaching methods |
| To inform parents/guardians during group meetings | To determine progress regarding school goals |
| To create group action plans | To inform other schools about an individual student by means of an educational report (school transition of the student) |
| To determine the group performance compared to the national performance | To give feedback to students in order to formulate their own learning goals |
| To inform people via the school prospectus or school website | To make decisions about students' transition year |
| To create school or annual plans | To inform colleagues about the student group during a group discussion or transmission |
| To determine the school's performance compared to the national performance | To place students into different groups for differentiation |
| To inform the individual student about his/her performance | To compare parallel groups regarding their progress |
| To create individual action plans for high performing students | To inform the inspectorate |
| To determine individual students' progress regarding learning goals or content | To create professional plans (performance appraisals, career decisions) |
| To inform the student group about their performance | To compare the performance of a student group with (those of) previous years. |

Table A.2

Questionnaire 2.

Respondents' background

1. What is your gender?
a) Male; b) Female
1. What is your sex?
3. In which district do your children attend school?
a) Friesland; b) Groningen; c) Drenthe; d) Overijssel; e) Gelderland; f) Limburg; g) Noord-Brabant; h) Zeeland; i) Utrecht; j) Noord-Holland; k) Zuid-Holland; l) Flevoland
4. What is your highest level of education?
a) No education/primary education; b) preparatory secondary vocational education; c) general secondary education; d) vocational education; e) senior general secondary education/university preparatory education f) univeristy of applied sciences; g) Master of Arts/Science/PhD
5. In which grade is your oldest child?
a) Lowest grades (Group 1,2); b) Middle grades (Group 3,4,5); c) Upper grades (Group 6,7,8); d) My oldest child has left primary school
6. Does your school use any of the following principles?
a) Anthroposophy; b) Dalton; c) Freinet; d) Jenaplan; e) Montessori; f) O4NT; g) none of these

Central questions

7. Which of the following purposes do you think best suits the reason to test your child at school?
a) Determining the level of your child; b) Adapting instruction to the educational needs of your child; c) Reporting and communicating the results to parents, the school board or inspectorate
8. Do you receive the score reports of your child from the student monitoring system?
a) No; b) Yes, in the score report of my child; c) Yes, during individual meetings with the teacher; d) Yes, during group meetings with parents; e) Other:
9. Would you like to support the learning process of your child?
a) No, in my opinion, this task belongs to the school; b) Yes, by means of the following actions . . .
10. This research looks at how test results are presented. What information from test results would you like to receive about your child? It is possible to give more than one answer.
a) The test scores of my child, focusing especially on the different subjects; b) The progress of my child with regard to the different subjects; c) The level at which my child is, compared to that of other children, with regard to the different subects; d) The level at which my child is with regard to the different parts of a subject; e) Learning material suggestions with regard to the different subjects in order to help my child in his/her learning; f) Other:

Appendix B.

Table B.1

Percentage of teachers, internal coaches and principals choosing an action as desired use and actual use.

Actions	Desired use			Actual use		
	Teachers n = 119	Internal coaches n = 27	Principals n = 12	Teachers n = 119	Internal coaches n = 27	Principals n = 12
Communicating learning						
To inform parents/guardians during individual meetings or by means of score reports	77.3	77.8	66.7	92.4	92.6	75.0
To inform the individual student about his/her performance	47.9	74.1	50.0	40.3	48.1	58.3
To inform other schools about an individual student by means of an educational report (school transition of the student)	52.1	77.8	50.0	60.5	77.8	75.0
To inform parents/guardians during group meetings						
To inform the student group about their performance	16.0	22.2	33.3	18.5	7.4	25.0
To inform colleagues about the student group during a group discussion or transmission	20.2	37.0	41.7	14.3	7.4	25.0
To inform people via the school prospectus or school website	61.3	77.8	50.0	73.9	70.4	50.0
To inform the school board or participation council	5.0	14.8	41.7	4.2	7.4	50.0
To inform the inspectorate	8.4	44.4	83.3	14.3	48.1	100.0
	20.2	55.6	58.3	46.2	77.8	100.0
Supporting learning						
To create individual action plans for low performing students	66.4	85.2	75.0	73.9	77.8	50.0
To create individual action plans for high performing students	63.0	81.5	58.3	53.8	63.0	41.7
To give feedback to students in order to formulate their own learning goals	45.4	77.8	50.0	16.0	29.6	33.3
To adapt instruction to educational needs	71.4	77.8	66.7	89.1	85.2	66.7
To place students into different groups for differentiation	57.1	85.2	83.3	50.4	77.8	83.3
To create school or annual plans	68.1	74.1	58.3	75.6	77.8	58.3
To formulate policy regarding the selection of a new teaching method	13.4	44.4	91.7	11.8	25.9	91.7
To create professional plans (performance appraisals, career decisions)	36.1	51.9	50.0	15.1	33.3	41.7
	9.2	18.5	33.3	5.9	3.7	50.0
Evaluating learning						
To determine the individual students' performance compared to the national performance	35.3	66.7	58.3	55.5	85.2	66.7
To determine individual students' progress regarding learning goals or content	66.4	77.8	58.3	73.9	85.2	58.3
To make decisions about students' transition year	37.8	55.6	33.3	48.7	63.0	58.3
To determine the group performance compared to the national performance	37.0	77.8	50.0	48.7	77.8	83.3
To determine group progress regarding learning goals or content	54.6	77.8	50.0	57.1	59.3	50.0
To compare parallel groups regarding their progress	15.1	44.4	25.0	14.3	29.6	25.0
To determine the school's performance compared to the national performance	26.1	66.7	75.0	20.2	74.1	91.7
To determine progress regarding school goals	30.3	74.1	50.0	32.8	66.7	75.0
To compare the performance of a student group with (those of) previous years.	33.6	66.7	58.3	49.6	77.8	75.0
No use of test results	5.0	8.3	5.1	1.7	0	0

References

AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Aschbacher, P. R., & Herman, J. L. (1991). *Guidelines for effective score reporting (CSE technical report 326)*. Los Angeles, CA: UCLA Center for Research on Evaluation, Standards and Student Testing (CRESST).

Baird, J. A., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the field review assessment and learning (case number 13/4697)*. Oslo: Knowledge Center for Education.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <http://dx.doi.org/10.1007/s11092-008-9068-5>.

Blok, H., Otter, M. E., & Roeleveld, J. (2001). *Het gebruik van leerlingvolgsystemen anno 2000 [The use of student monitoring systems in the year 2000]*. Amsterdam, the Netherlands: SCO-Kohnstam Instituut.

Bradshaw, J., & Wheeler, R. (2009). *National foundation for educational research: International survey of results reporting (OFQUAL 10/4705)*. London: Office of Qualifications and Examinations.

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, NJ: Pearson Education.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston: Harvard Business School Press.

Ebbeler, J., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2016). Effects of a data use intervention on educators' use of knowledge and skills. *Studies in Educational Evaluation*, 48, 19–31. <http://dx.doi.org/10.1016/j.stueduc.2015.11.002>.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17(2), 145–220. <http://dx.doi.org/10.1207/s15324818ame1702>.

Gummer, E., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, 117(4), 1–22 Retrieved from <https://www.tcrecord.org>.

Hambleton, R. K., & Slater, S. C. (1997). *Are NAEP executive summary reports understandable to policy makers and educators? (CSE Technical Report 430)*. Los Angeles, CA: UCLA Center for Research on Evaluation, Standards and Student Testing (CRESST).

Hambleton, R. K., & Zenisky, A. L. (2013). Reporting test scores in more meaningful ways: A research-based approach to score report design. In K. E. Geisinger (Ed.), *Handbook of testing and assessment in psychology* (pp. 479–494). Washington, DC: APA.

Hattie, J. A. C., & Brown, G. T. L. (2008). Technology for school-based assessment and assessment for learning: Development principles from New Zealand. *Journal of Educational Technology Systems*, 36(2), 189–201. <http://dx.doi.org/10.2190/et.36.2.g>.

Hattie, J. (2009). Visibly learning from reports: The validity of score reports. *Paper presented at the meeting of the National Council on Measurement in Education*.

Hellrung, K., & Hartig, J. (2013). Understanding and using feedback: A review of empirical studies concerning feedback from external evaluations to teachers. *Educational Research Review*, 9, 174–190. <http://dx.doi.org/10.1016/j.edurev.2012.09.001>.

Jaeger, R. M. (1998). *NAEP validity studies: Reporting the results of the national assessment of educational progress (Working paper 2003-11)*. Palo Alto, CA: American Institutes for Research.

Ledoux, G., Blok, H., Boogaard, M., & Krüger, M. (2009). *Opbrengstgericht werken: Over de waarde van meetgestuurd onderwijs [Data-driven decision making: About the value of measurement oriented education]*. Amsterdam, the Netherlands: SCO-Kohnstam Instituut.

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23(2), 26–32. <http://dx.doi.org/10.1111/j.1745-3992.2004.tb00156.x>.

Mandinach, E. B., & Jackson, S. S. (2012). *Transforming teaching and learning through data-driven decision making*. Thousand Oaks, CA: Corwin.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71–85. <http://dx.doi.org/10.1080/00461520.2012.667064>.

- Marsh, J. A. (2012). *Interventions promoting educators' use of data: Research insights and gaps*, 114 (11), 1–48 Retrieved from <http://www.rdc.udel.edu>.
- Meijer, J., Ledoux, G., & Elshof, D. (2011). *Gebruikersvriendelijke leerlingvolgsystemen in het primair onderwijs [User-friendly pupil monitoring systems in primary education] (Report 849)*. Amsterdam: Kohnstamm Instituut.
- Monaghan, W. (2006). *The facts about subscores (Report No. RDC-04)*. Princeton, NJ: Educational Testing Service.
- NEGP (1998). *Talking about tests: An idea book for state leaders*. Washington, DC: U.S. Government Printing Office.
- Newby, P. (2010). *Research methods for education*. Harlow, England: Longman.
- Newton, P. E. (2005). The public understanding measurement inaccuracy. *British Education Research Journal*, 31(4), 419–442. <http://dx.doi.org/10.1080/01411920500148648>.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11. <http://dx.doi.org/10.1080/00405840802577536>.
- Ryan, J. M. (2006). Practices, issues, and trends in student test score reporting. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 677–710). Mahwah, NJ: Lawrence Erlbaum.
- Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26(3), 482–496. <http://dx.doi.org/10.1016/j.tate.2009.06.007>.
- Schildkamp, K., & Lai, M. K. (2013). Data-based decision making: An overview. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 9–21). Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-007-4816-3_1.
- Schildkamp, K., & Teddlie, C. (2008). School performance feedback systems in the USA and in The Netherlands: A comparison. *Educational Research and Evaluation*, 14(3), 255–282. <http://dx.doi.org/10.1080/13803610802048874>.
- Van Geel, M., Keuning, T., Visscher, A. J., & Fox, J. (2016). Changes in educators' data literacy during a data-based decision making intervention. In T. Keuning, & M. J. M. Van Geel (Eds.), *Implementation and effects of a schoolwide data-based decision making intervention: A large skill study* (pp. 73–96). Universiteit Twente: Enschede, The Netherlands.
- Van der Kleij, F. M., & Eggen, H. J. M. (2013). Interpretation of the score reports from the computer program LOVS by teachers, internal support teachers and principals. *Studies in Educational Evaluation*, 39(3), 144–152. <http://dx.doi.org/10.1016/j.stueduc.2013.04.002>.
- Van der Kleij, F. M., Eggen, T. J. H. M., & Engelen, R. J. H. (2014). Towards valid score reports in the computer program LOVS: A redesign study. *Studies in Educational Evaluation*, 43, 24–39. <http://dx.doi.org/10.1016/j.stueduc.2014.04.004>.
- Vanhoof, J., Verhaeghe, G., Verhaeghe, J. P., Valcke, M., & Van Petegem, P. (2011). The influence of competences and support on school performance feedback use. *Educational Studies*, 37(2), 141–154. <http://dx.doi.org/10.1080/03055698.2010.482771>.
- Verhaeghe, G., Vanhoof, J., Valcke, M., & Van Petegem, P. (2011). Effecten van ondersteuning bij schoolfeedbackgebruik [Effects of support in school feedback use]. *Pedagogische Studien*, 88(2), 90–106.
- Wainer, H., Hambleton, R. K., & Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement*, 36(4), 301–335. <http://dx.doi.org/10.1111/j.1745-3984.1999.tb00559.x>.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Erlbaum.
- Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- Zapata-Rivera, J. D., & Katz, I. R. (2014). Keeping your audience in mind: Applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice*, 21(4), 442–463. <http://dx.doi.org/10.1080/0969594X.2014.936357>.
- Zapata-Rivera, D., & Zwick, R. (2011). *Test score reporting: Perspectives from the ETS score reporting conference (ETS RR-11-45)*. Princeton, NJ.
- Zapata-Rivera, D., VanWinkle, W., & Zwick, R. (2012). *Applying score design principles in the design of score reports for CBAL™ teachers (ETS RM-12-20)*. Princeton, NJ.
- Zenisky, A. L., & Hambleton, R. K. (2012). Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice*, 31(2), 21–26. <http://dx.doi.org/10.1111/j.1745-3992.2012.00231.x>.
- Zwick, R., Sklar, J. C., Wakefield, G., Hamilton, C., Norman, A., & Folsom, D. (2008). Instructional tools in educational measurement and statistics (ITEMS) for school personnel: Evaluation of three web-based training modules. *Educational Measurement: Issues and Practice*, 27(2), 14–27. <http://dx.doi.org/10.1111/j.1745-3992.2008.00119.x>.
- Dorien Hopster-den Otter** is a Ph.D. candidate at the Research Center for Examinations and Certification (RCEC), a collaboration between Cito and the University of Twente. Her research interests include formative assessments, psychometrics and score report development. She is currently working on projects relating to the quality of educational tests and the development of an educational master's programme for test development.
- Saskia Wools** is an educational researcher and manager of prototyping at Cito. Her research interests include the validity and validation of educational assessments. She is currently working on innovative projects involving formative assessments, educational technology and assessment quality.
- Theo J. H. M. Eggen** is a senior research scientist at the Psychometric Research Center of Cito and a professor of psychometrics at the University of Twente in the Netherlands. He is director of the Research Center for Examinations and Certification (RCEC). His research interests include the quality of educational testing and computerized (adaptive) testing. He is also working on projects like PISA (international educational survey) and on a variety of projects on formative educational assessment.
- Bernard P. Veldkamp** is head of the Department of Research Methodology, Measurement and Data Analysis and the scientific director of the Research Center for Examination and Certification (RCEC). His research interests include measurement optimization and behavioral data science. His interests focus on computerized assessment. His current projects include obtrusive assessment in online learning and a review study on the conditions for effective formative assessment.