Studies in Educational Evaluation 52 (2017) 12–23

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 24 November 2016

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.

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

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 inspectorate[1] 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 & Wheater, 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