Studies in Educational Evaluation 42 (2014) 79–90

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

The effects of professional development on the attitudes, knowledge and skills for data-driven decision making

L. Staman, A.J. Visscher *, H. Luyten

University of Twente, Department of Educational Sciences, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 4 April 2013
Received in revised form 12 November 2013
Accepted 18 November 2013
Available online 20 December 2013

Keywords:
Data-driven decision making
Data-driven teaching
Student monitoring system
Professionalization
Data use

ABSTRACT

The Dutch government and School Inspectorate encourage schools to use the student performance data they can obtain from their student monitoring systems to maximize student performance in a systematic and goal-oriented way. Research by the same Inspectorate (Inspectie van het Onderwijs, 2010) shows that students in schools which do so outperform students in schools where data-driven decision making (DDDM) is as yet less developed. The University of Twente developed a training course in which school teams learn to utilize data from computerized student monitoring systems in order to improve instructional quality and student performance. Parallel to the training activities, training effects are studied. The research findings show that the training activities had a positive effect on school staff's DDDM knowledge and DDDM skills. Staff attitudes towards DDDM were already high on the pre-tests and remained high on the post-tests.

Introduction to data use and its prerequisites

What is meant by "data use"?

A broad definition of data-driven decision making (DDDM) could be 'the process of using data to enhance education' (Schuyler Ikemoto & Marsh, 2008, chap. 5). This definition is indeed broad as it refers to *various kinds of data*, e.g. data on teachers (their performance, absence rates, etc.), students (their test scores, school careers, drop-out rates, absence rates, etc.), the school as a whole (e.g. average drop-out rates, and percentages of class repeaters, relationships between students' scores on final examinations and on school internal examinations, etc.), and it could even include data about a feature such as a 'timetable' (e.g. timetable quality from an educational point of view), and other relevant features within schools. *Relationships* between various types of data can be productive to explore in the improvement of educational quality, e.g. when do students play truant most (which lessons, subjects, teachers?), and, does a relationship exist between teacher and student absenteeism on the one hand and their performance on the other?

The broad DDDM definition not only include a variety of data but also various types of decisions related to the improvement of educational quality in terms of *the level* at which those decisions are taken, e.g. classroom, school, and school board. Moreover, there will be variation in terms of *the stages of decision making* intended to improve the quality of schooling (Visscher & Ehren, 2011) (see Fig. 1):

- evaluating the functioning of a school board/school/classroom (what is our level of performance? and what needs improvement?) based on the analysis of available data;
- setting performance goals what do we want to improve?;
- deciding how to attempt to/what strategy to use to accomplish the goals set, and, finally, implementing the strategy for goal accomplishment (and monitoring its execution).

Given the enormous variety of concepts potentially included in the term 'data use' in DDDM, it is advisable/desirable to specify precisely the type of data and decision making referred to. Too often the DDDM concept is used in too general a way. In the view of the authors the most important data and decisions within school systems relate directly to what happens and will happen in classrooms. Effective schools add much to the school entrance levels of their students, and research has repeatedly shown that the most significant cause of variation in students' added value is found in classroom characteristics (Kane & Staiger, 2005; Nye, Konstantopoulos, & Hedges, 2004; Scheerens & Bosker, 1997). In other words, the quality of classrooms and teachers vary considerably, which means that there is also considerable room for improvement at that level. Knowing that the behaviour of

* Corresponding author. Tel.: +31 53 4893609.
E-mail addresses: l.staman@utwente.nl (L. Staman), a.j.visscher@utwente.nl (A.J. Visscher), j.w.luyten@utwente.nl (H. Luyten).

0191-491X/$ – see front matter
http://dx.doi.org/10.1016/j.stueduc.2013.11.002

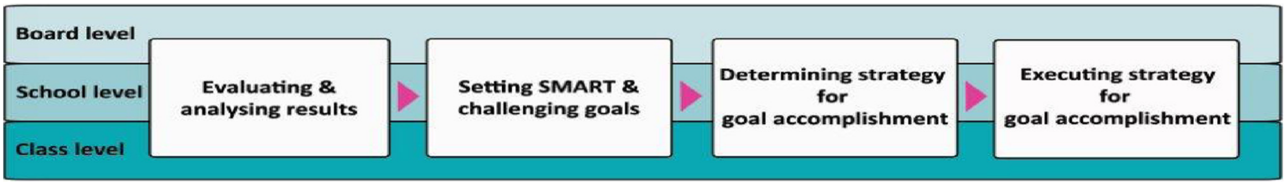

Fig. 1. Decision-making levels and stages.
Source: Keuning and Van Geel (2012).

teachers is the most powerful malleable factor in maximizing teaching quality and student achievement implies that quality data on classroom processes and results form a very important lever for the improvement of classroom and school performance. It seems logical that this is where DDDM can impact on the effectiveness of schools.

Features of the CITO standardized tests for primary school students

The focus of this article is on the use of the results of the standardized tests taken by primary school students for the improvement of school quality. These tests have been developed by the Dutch CITO (Central Institute for Test Development) and approximately 90% of Dutch primary schools use one or more components of the CITO student monitoring system (Ledoux, Blok, Boogaard, & Krüger, 2009). Almost all Dutch primary schools either use the CITO software to analyze the test results, or a web-based data-analysis system developed by a commercial vendor: ESIS, or ParnasSys.

Using the student monitoring system (SMS) student achievement can be monitored on a longitudinal basis across all primary school grades, at the level of individual students, grades, student cohorts, and the school as a whole.

The standardized tests are taken once or twice a school year. The frequency varies between subjects. When the tests are taken twice a year, this is usually once mid-year and once at the end of the year. The system supports the longitudinal analysis of performance data, i.e. performance across time on the various tests taken.

The psychometric characteristics of the CITO tests allow the results of the various tests (e.g. those taken in Grade 2 and those taken in Grade 6) to be represented on the same ability scale. As a result student **growth** between tests can be determined. Student growth between the E3 test (taken at the end of Grade 3) and the M4 test (the test taken half way through Grade 4) is shown in Fig. 2 for A-students (the top 25% of students), and for B-, C-, D-, and E-students (the 10% of students performing at the bottom of the performance distribution). Fig. 2 also shows national average growth for the 'average' student' (the white line in Fig. 2). National average growth between E3 and M4 has been found to be 12 points. A students' performance between E3 and M4 thus can be evaluated by comparing his/her growth with these 12 points of **average**, 'normal' growth.

The CITO tests also enable the utilization of other important information on the performance of students. Fig. 3, for example, shows how the students in class 6A, 6B, 7A and 7B have performed in comparison with the national average (average performance is represented by the dotted line; in an average class 50% of students performs above the thick dotted line, and 50% performs below). The colours within each rectangle show the percentage of A-students (the 25% top students), B-, C-, D- and E-students in a class. Class 7B has a high number of D-students. On the basis of this information a school manager can instantly see how each class in the school has performed.

Fig. 4 shows how four student year groups (year groups 4, 5, 6 and 7) performed over the four school years from 2005–2006 to 2008–2009. The performance of a student group is depicted on the basis of the E-test taken at the end of a school year. If the vertical line is lower than the M-line, then performance is *below* average national performance. Where it is above the M-line (as is the case for all data collection points in this figure) it is *above* the average national performance. If a year group were found to be frequently underperforming, this information would be very valuable for school managers who probably would try to find the cause of such structural underperformance, and try to solve the problem.

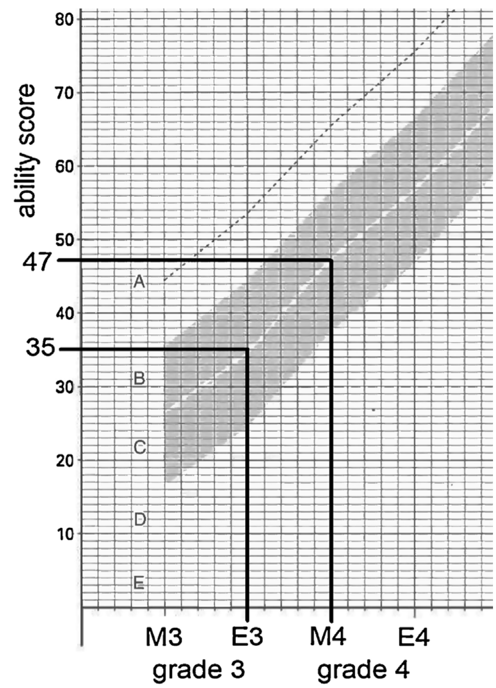

Fig. 2. Student performance growth between two tests.

Fig. 5 shows performance *growth* between two tests. Each rectangle represents a student. The figure shows how much an individual student has grown between two tests and whether this growth is above or below the national average growth (the top dotted line), and above or below the average growth in his/her class (the lower dotted line). The student at the right under the 0-line has not shown any growth on these tests!

Computation of the added value between two or more measurement moments is important as a student may perform at a relatively high level in a test but there may have been little growth since the previous test. Conversely it is useful to know if a lower performing student has grown more than the average since the previous test.

Fig. 6 is valuable for teachers as it shows to what extent students master the various *categories* within the arithmetic subject, and in which categories students deviate from their overall performance pattern. For example, the discrepancies in the profile of results of student Lejla van Motmar who is a D-student with a Total Ability Score of 43 are somewhat striking as she performs quite well on most arithmetic categories, but her score of 12 for