University of Groningen

Effects of a Data-Based Decision Making Intervention on Student Achievement

Staman, Laura; Timmermans, Anneke; Visscher, A. J.

Published in:
Studies in Educational Evaluation

DOI:
[10.1016/j.stueduc.2017.07.002](https://doi.org/10.1016/j.stueduc.2017.07.002)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Staman, L., Timmermans, A., & Visscher, A. J. (2017). Effects of a Data-Based Decision Making Intervention on Student Achievement. *Studies in Educational Evaluation*, 55, 58-67.
<https://doi.org/10.1016/j.stueduc.2017.07.002>

Copyright

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

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Effects of a data-based decision making intervention on student achievement

L. (Laura) Staman^{a,*}, A.C. (Anneke) Timmermans^b, A.J. (Adrie) Visscher^a

^a University of Twente, The Netherlands

^b University of Groningen, GION Education/Research, The Netherlands

ARTICLE INFO

Keywords:

Data-based decision making

Student achievement

Professional development

Intervention

ABSTRACT

Data-based decision making (DBDM) is becoming important for teachers due to increasing amounts of digital feedback on student performance. In the quasi-experimental study reported here, teachers, principals, and academic coaches from 42 schools were trained for two years in using the results of half-year interim assessments for providing students with tailor-made instruction. Our results did not show any main effects of this DBDM training trajectory on student achievement but did indicate interaction effects with students' low prior achievement levels and socioeconomic status. Teachers experience difficulties in translating student progress data into adaptive instruction in the classroom. Implications of our findings for teacher professionalization are discussed.

1. Introduction

Data-Based Decision Making (DBDM) is growing internationally. Since the 'No Child Left Behind' (2001) act was introduced in the United States American schools are expected to make decisions based on (test) data, and to meet the standards for 'Adequate Yearly Progress' (Carlson, Borman, & Robinson, 2011). Several years ago, the Dutch government developed a DBDM policy because they were concerned about the performance level of Dutch students as indicated by international comparative studies (Ministry of Education, 2007). The results of the TIMSS, PIRLS, and PISA surveys indicated, according to the Dutch government, that the performance of Dutch students was declining. The idea was that teachers and schools should make (more) use of the results from interim assessments to provide instruction matching with the, varying, instructional needs of their students (Dutch). The assumption behind this policy was that implementing DBDM could maximize the performance of all students as students were expected to be taught more in line with their needs (as indicated by the interim assessments) (Visscher, Peters, & Staman, 2010). Educators have access to an increasing amount of data.

Ikemoto and Marsh (2007) define DBDM as follows: "teachers, principals, and administrators systematically collecting and analyzing data to guide a range of decisions to help improve the success of students and schools" (p. 108). According to Dunn, Airola, Lo, and Garrison (2013) DBDM at classroom level entails "the identification of patterns of performance that unveil students' strengths and weaknesses relative to students' learning goals as well as the selection and planning of instructional strategies and interventions to facilitate student

achievement of learning goals" (p. 225). This could enhance student learning if teachers reflect (more) on the impact of the instruction they have given, and attempt to provide tailor-made instruction based on their knowledge of how students perform and which learning problems they face (Schildkamp & Kuiper, 2010). Effective DBDM presupposes that teachers learn to utilize the student performance information they can retrieve from their student monitoring systems, and that they learn to choose and implement effective instructional approaches that may help improve students' mastery of subject matter content better, see Fig. 1.

However, applying DBDM effectively may be cumbersome, as previous research has shown that teachers struggle with several aspects of DBDM. Teachers struggle with utilizing test data and other data about students effectively (e.g., Marsh, Pane & Hamilton, 2006; McCaffrey & Hamilton, 2007). DBDM also presupposes that teachers master complex didactical skills, such as the differentiation of instruction in response to observed performance differences between students. The Dutch Inspectorate of Education points to the fact that teachers experience difficulties with executing these complex differentiation skills (Inspectorate of Education, 2013). Teachers need support in implementing DBDM fully (Slavin, Cheung, Holmes, Madden, & Chamberlain, 2013). We noticed that generally limited attention is paid to equipping teachers for going from data to tailor-made instruction. It was our expectation that paying attention to all DBDM components and to the prerequisites for effective DBDM might make DBDM more effective.

The University of Twente developed a training trajectory in which school teams learn to work systematically in a DBDM manner for the

* Corresponding author at: University of Twente, Faculty of BMS, Department of Teacher Development, P.O. Box 217, 7500AE Enschede, The Netherlands.
E-mail address: lstaman@rocvantwente.nl (L.L. Staman).

Fig. 1. Levels and components of DBDM (Keuning & Van Geel, 2012).

subject mathematics: the Focus project. In the Focus project schools were trained for working with the student test data in a wide sense: data interpretation and translating the data into tailor-made instructional strategies. The goal of the study reported here was to evaluate whether the Focus intervention that was meant to equip teachers and schools for DBDM was effective in terms of accomplishing higher levels of student achievement than a control group which was not trained for DBDM. In the following overview we first will present the results of multiple studies on the effects of DBDM interventions, and then describe our quasi-experimental study that was done to investigate the effects of the Focus DBDM intervention. In the latter we also describe how the Focus intervention was meant to equip teachers with the attitudes, knowledge and skills to adequately apply DBDM.

1.1. DBDM effects on student performance

Using student performance data (often in combination with other relevant data) to improve student performance receives much attention from scholars around the world. Various terms are used for this approach: data-driven teaching (e.g., Fullan, 2007), data-driven decision making (e.g., Wohlstetter, Datnow, & Park, 2008), and data-based decision making (e.g., Carlson et al., 2011). Despite the differing terminology, the approach is similar: to systematically work on educational improvement and basing improvement-oriented measures on student performance data. Student performance data can inform teachers about students' individual instructional needs, which can be the basis for tailor-made instruction. Data that is useful for DBDM practice includes for example the results of standardized tests, curriculum-based tests and other types of student classroom work (e.g., Lai & Schildkamp, 2013; Mandinach & Gummer, 2013).

Despite high expectations of DBDM in terms of improved student performance, scientific evidence on the effectiveness of DBDM interventions is still rather limited, meaning that current recommendations regarding the *implementation* and the *effects* of DBDM are not founded on a strong empirical basis (e.g., Hamilton et al., 2009). DBDM interventions have been evaluated in a number of (quasi-) experimental studies. Most research into the effects of DBDM reports the findings of small scale studies and non-experimental designs; we need stronger research designs. Research into the links between the DBDM components (see Fig. 1) is also scarce, or, if such studies are done, the effect of DBDM on student achievement is not investigated (Lai & Schildkamp, 2013; Marsh, 2012).

An example of one of the rare experimental studies into the effects of a DBDM intervention is the study by Carlson et al. (2011) that included coaching schools in analyzing and interpreting student performance, and the support of schools in selecting and implementing evidence-based improvement strategies. The results of this study indicated that students from schools in which student progress was analyzed using benchmark assessments performed significantly better in mathematics than a control group. The difference between both conditions proved to be 0.06 standard deviations at the student level, 0.20 at the school level, and 0.21 at the district level.

The same DBDM intervention was also investigated experimentally by Slavin et al. (2013), however, in this case not over the course of one, but four years. Both, statistically significant positive as well as statistically non-significant effects were found for reading and mathematics

in grades five and grade eight. In grade five, statistically significant differences for reading and mathematics were found in the third and fourth intervention year with effect sizes varying between 0.24 and 0.50. In grade eight, positive statistically significant effects were found for mathematics and reading during the first intervention year. In the second intervention year, positive significant effects were only found for reading, and in the fourth intervention year only for mathematics. The effect sizes varied between 0.17 and 0.31. An important finding is that the largest effects were found after four intervention years. According to Slavin et al. (2013), schools started to change and improve instruction only after the third intervention year. The first two years of the intervention mainly included the analysis and interpretation of available data; "...first and second year interventions were analogous to taking a patients temperature, not providing a treatment" (Slavin et al., 2013, p. 390). Duration is an important characteristic of effective professional development, because learning and changing practice takes time. Changing teacher behavior is not easy, because of the many obligations face in their work (Van Veen, Zwart, & Meirink, 2010). In the view of Supovitz and Turner (2000) it takes eighty hours to change teacher behavior, according to Timperley (2008) it takes one to two.

Ledoux, Blok, Bogaard, and Krüger (2009) stated that Dutch research into DBDM and its effects on student performance is still limited. An example of a Dutch study into DBDM is the study by Van Kuijk (2014). An effect of 0.37 for reading comprehension was reported in a study using a matched-pairs design (Van Kuijk, 2014), whereas the same intervention did not lead to statistically significant effects on students' mathematics performance (Ritzema, 2015). In both projects, teachers were trained in formulating standards and performance goals for each student as well as in utilizing data and learning how to apply modelling for reading and mathematics instruction.

Faber and Visscher (2014) conducted a meta-analysis to summarize the results of experimental studies on the use of digital student monitoring systems, an important DBDM component (Visscher et al., 2010), on student performance. An average effect size of 0.38 was found for interventions aimed at the use of digital student monitoring systems with small groups of children. For the implementation of digital student monitoring systems in larger contexts (school-wide, or school board-wide) a positive, but smaller effect was found (effect size 0.06). Factors that proved to promote the intended effects of digital student monitoring systems on student performance are a high feedback frequency, and the use of monitoring systems that include instructional advice in their feedback to practitioners. Finally, interventions involving system implementation activities (e.g., training teachers for system use) at least once a month proved to be more effective.

Overall, the results of the studies presented here show a mixed picture of the effects of DBDM on students' (mathematics) performance. In some studies, the DBDM intervention effects are significant, whereas in others they are not. Also, the significant effects' magnitude varies considerably. These discrepant findings may be due to differences in the feedback that teachers receive in terms of frequency and content (e.g., with, or without follow-up advice). Differences between DBDM interventions are also important, e.g., how and how much users are trained in system use in the broad sense of the term: learning to use data to provide instruction in line with observed student needs.

Fig. 2. Overview of the Focus-meetings during the academic year 2010–2011.

1.2. Research questions

Given the Dutch DBDM policy to encourage schools to use assessment data for tailor-made instruction, we decided to support school teams for DBDM. Teachers were coached to learn to analyze the data they could retrieve from their student monitoring systems; to set the performance goals they want to accomplish with their students, and to design and to implement instructional approaches for accomplishing the goals set. The previous led to the following research question: What is the effect of participating in the Focus intervention on students' performance in mathematics?

2. Method

2.1. The focus intervention

The Focus intervention rationale is presented in Fig. 1. School teams were trained and coached in evaluating and analyzing their student monitoring system data, setting SMART (Specific, Measurable, Attainable, Realistic, and Time-bound) and challenging performance and subject matter content goals, to accomplish with their students, in formulating strategies to accomplish the goals, and in executing the planned strategies. This is an approach that can be applied at the classroom level, the school level, and at higher administrative levels (e.g., the district level, or the school board level). The Focus intervention studied in this project was mainly aimed at DBDM at the classroom level. During the intervention, teachers from participating schools learned how to carry out each of the four DBDM components in Fig. 1. The various elements of this cycle were elaborated in a protocol. This protocol was a kind of road map used to analyze the student monitoring data step by step. The protocol was presented and adapted during the training sessions (schools adapted them to their own context). Schools used this protocol twice per academic year (halfway through the academic year in January/February and at the end of the academic year in June/July).

All members of the school team participated in the intervention. This is important for effective professional development (Timperley, 2008; ; Van Veen et al., 2012) because colleagues then can communicate and collaborate. The Focus intervention lasted two years and teachers from pre-kindergarten up to grade six (the final grade of Dutch primary education in which students are 11/12 years old), the principal, and the academic coach participated in the training trajectory. The principal and the academic coach were trained both in intervention years, while teachers participated only one year. Teachers were split into two groups. One group of teachers (pre-kindergarten up to grade three) received the Focus training in the academic year 2010–2011. In this academic year, the academic coaches and the principals were trained together with the teachers from pre-kindergarten up to grade three. The teachers from grades four up to grade six received the training in the academic year 2011–2012. This group of teachers was trained separately from the academic coaches and the principals (who were trained in meetings P1, P2, and P3 in Fig. 3). Training them together would have meant too much repetition for the academic coaches

and principals. The coaches and principals followed a separate program about DBDM at the school level.

Some schools chose not to participate with their (pre-) kindergarten teachers in the intervention because they considered it inappropriate to work with student assessment data in the early years of schooling. To promote cooperation between and within schools, participating schools were allocated to one of eight training groups. A training group included between 25 and 40 participants. All training groups were trained by the same educational consultant who specialized herself in DBDM and who worked together with the researchers who had developed the intervention. She studied educational sciences and also was a teacher herself.

Fig. 2 show how the Focus meetings were distributed over the academic year 2010–2011; the content of each of the seven meetings can be found in the Appendix A. Schools were first taught to analyze student performance data with their student monitoring system. Next, they learned how to set student performance goals given what they had learned about their students' performance. In the following step, they were coached to design instructional approaches that were expected to be effective in accomplishing their goals. In the final step, teachers had to implement the tailor-made instructional plan in the classroom. After completing all the steps, teachers had to analyze students' progress (after half a year, when students took the next standardized test) and thereafter go through all the steps again.

Fig. 3 presents the distribution of the Focus meetings over the academic year 2011–2012.

2.2. Participants

Eighty-four primary schools participated in this quasi-experimental study, 42 schools in the experimental and 42 in the control condition. The experimental schools were asked to participate after presentations had been given about the Focus intervention. The control schools were selected by means of propensity score matching. The following school background characteristics were used as matching variables: denomination, the percentage of students with a specific socio-economic status, the percentage of ethnic students, schools' type of inspection supervision (intensive or less intensive, reflecting the performance level of the school), the percentage of female staff, and the school size. Using nearest neighbor propensity score matching, a control school was selected for each of the experimental schools. So we composed an experimental and control group that are as much as possible similar on background characteristics. In that way a student performance difference between both groups can be attributed to the fact that one group received the data use intervention whereas the other not. A 1:1 matching method without replacement was applied; as such schools could only be part of one pair. Table 1 presents the descriptive statistics for both groups of schools.

Student achievement data and student background characteristics were collected by exporting them from the student monitoring systems of participating schools. Mathematics achievement data from the academic years 2008–2009, 2009–2010, 2010–2011, and 2011–2012 were collected. These data reflect the two years before the start of the Focus

Fig. 3. Overview of the Focus meetings during the academic year 2011–2012.

Table 1
Descriptives for experimental and (potential) control schools.

	Experimental schools (n = 42)	Potential control schools (n = 7267)	Selected control schools (n = 42)
	n (%)	n (%)	n (%)
Denomination			
Protestant-Christian	11 (26.2)	2202 (30.3)	16 (38.1)
Catholic	17 (40.5)	2126 (29.3)	14 (33.3)
Public	13 (31.0)	2468 (34.0)	12 (28.6)
Other	1 (2.4)	471 (6.5)	0 (0.0)
Inspection school supervision regime			
Regular	41 (97.6)	6604 (90.9)	41 (97.6)
For weak schools	1 (2.4)	430 (5.9)	1 (2.4)
For very weak schools	1 (2.4)	68 (0.9)	0 (0.0)
Unknown	0 (0.0)	165 (2.3)	0 (0.0)
Percentage SES	Average 22	Average 15	Average 20
Percentage minority students	11	11	7
Percentage female staff	83	82	86
School size ^a	14.85	15.02	11.40

^a As measured in FTE (fulltime-equivalent).

intervention and the two years in which the Focus intervention took place.

We could not analyze the effects of the Focus intervention for all students in participating schools. Students' pre-kindergarten and kindergarten scores could not be included in the analyses, because the scores of specific kindergarten tests could not be compared with the test scores in grade 1 and higher. Furthermore, mathematics tests were rarely administered in grade six, because students then had to take the 'School leavers' test (the final test administered at the end of primary education). The analyses therefore included the development of students' mathematics achievement from grade one up to grade five.

2.3. Instruments and variables

2.3.1. Mathematics achievement

Student achievement in mathematics was measured by means of students' test scores on standardized student tests as developed by Cito (the Dutch Testing Institution). By means of these tests, students' mathematics ability can be monitored throughout the entire primary school period. The tests have good overall reliability in all grades (Cronbach's α of at least 0.91; Janssen, Verhelst, Engelen, & Scheltens, 2010). The standardized mathematics tests are administered twice an academic year: halfway through (January) and at the end (June) of the academic year. The mathematics tests scores of the different grades have been calibrated to a single underlying latent scale by means of Item Response Models; more specifically the One Parameter Logistic Model was used (Janssen et al., 2010). The scores used in the current analyses are those that have been converted to this single scale (ranging from 0 to 169) and in which item difficulty has been taken into account. The schools participating in our study used both new and older versions of the mathematics tests. Student scores for the old test versions for grade one up to grade five could be rescaled by means of a table provided by Cito, to allow for the comparison of performance data across both test versions. The correlation of the latent scores from the old and the new test versions are high (Hop & Engelen, 2017).

2.3.2. Condition

A dummy variable indicating whether a student attended an experimental school (1), or a control school (0).

2.3.3. Student background

Two student background variables were derived from the student monitoring system: students' gender and their socio-economic background. The socio-economic status was measured on three ordinal categories based on the level of parental education.¹ The following three categories were distinguished: 1) average and high parental education (0.0), low parental education – both parents attended lower vocational education as their highest level of education (0.3), and very low parental education – the highest educational level of at least one of the parents was primary education (1.2).

2.4. Data analysis strategy

In order to estimate the effects of the Focus intervention on students' mathematics performance two different approaches were used, to compare the experimental and the control condition.

2.4.1. The cohort approach

The first approach was a longitudinal approach in which two student cohorts were studied. For each student, longitudinal data was available for four subsequent academic years. The data for these four academic years were used to study the development of students' mathematics achievement before and during the Focus intervention. For each student, a maximum of eight test scores could be used in this approach.

The first cohort included students who were in grade one in academic year 2008–2009. Teachers of cohort one students participated in the Focus training in the academic year 2010–2011 (grade 3 teachers) and in the academic year 2011–2012 (grade four teachers). This cohort included a total of 1984 students and 14,497 test scores. This implies that, on average, 7.3 test scores per student were available.

The second cohort included students who were in grade two in the academic year 2008–2009. In the case of students from the second cohort, the principal and the academic coach already participated in the Focus intervention in 2010–2011 but their teacher then did not yet. In 2011–2012, their teachers (grade four) also participated in the intervention. The second cohort included 1976 students and a total of 15,089 test scores; thus, on average 7.6 test scores per student were available. Table 3 presents the characteristics of the students in the experimental and the control group from the two cohorts.

By means of multilevel growth curve models, it was determined whether the experimental and the control condition differed from each other in terms of student performance growth. The MLwiN 2.28 software was used for this purpose (Rasbash, Steele, Browne, & Goldstein, 2009). The estimated multilevel growth curve models had three hierarchical levels: measurements of student achievement (level 1) nested within students (level 2), nested within schools (level 3). The development of students' mathematics achievement was modelled as a quadratic function of time, which enabled the detection of a potentially non-linear development (Table 2).

A series of growth curve models has been analyzed. In all models, students' mathematics scores at various measurement moments served as the dependent variable. In Model 1, no explanatory variables were included (empty model). In Model 2, time (representing the number of months from the start of the Focus intervention) was added to estimate the general trend in mathematics achievement. In Model 3, a number of variables and interactions were added to the model. In this model, four effects were tested. First, 'condition' was added as predictor variable, to test whether there is an intercept difference between the experimental and the control condition. Second, an interaction between time and the

¹ Dutch students are assigned extra 'weight' if their parents have a lower educational background. Students can obtain an extra weight of 0.3 or 1.2. Schools receive additional funding based on these extra student weights as it is assumed that schools with students with extra student weight have a more difficult job to do.

Table 2
Characteristics of the first and the second student cohort.

	Cohort 1		Cohort 2	
	Experimental n (%)	Control n (%)	Experimental n (%)	Control n (%)
Gender				
Boys	515 (48.3)	488 (53.2)	543 (50.6)	440 (48.1)
Girls	551 (51.6)	429 (46.8)	530 (49.4)	463 (51.3)
Unknown	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)
Socio-economic status				
0.0	781 (73.3)	585 (64.2)	816 (76.3)	574 (63.9)
0.3	100 (9.4)	71 (7.8)	95 (8.8)	66 (7.3)
1.2	80 (7.5)	27 (3.0)	60 (5.6)	40 (4.4)
Unknown	105 (9.8)	228 (25.0)	99 (9.3)	218 (24.3)
Total	1067 (53.8)	917 (46.2)	1073 (54.3)	903 (45.7)

experimental condition was included to test whether the experimental and the control condition showed similar trends before the start of the intervention. Having used the propensity score matching strategy, we expected no significant intercept and trend differences between experimental and control schools before the start of the intervention. Third, an interaction between time and the start of the intervention was included to assess changes in the development of students in the control group at the start of the intervention. Because no intervention took place in the control schools, we did not expect differences in the development of mathematics achievement before and after the start of the intervention for the students in the control group. Fourth, a three-way interaction was included to test whether the control and the experimental conditions showed different trends after the start of the intervention. We expected a more positive development of mathematics achievement for students in the experimental relative to the control group due to the Focus DBDM intervention. This was assessed with one-sided significance tests using $\alpha = 0.05$.

2.4.2. The grade approach

The second approach for comparing experimental and control schools was by analyzing data per *grade* in a pre-test post-test analysis. The outcomes of students in grades two and three at the end of academic year 2010–2011 (post-test) were compared to their achievement at the end of the previous academic year 2009–2010 (pre-test). The teachers in these grades participated in the Focus intervention in the academic year 2010–2011. Similarly for grades four and grade five, students' outcomes at the end of the academic year 2011–2012 (post-test) were compared with their outcomes at the end of the academic year 2010–2011 (pre-test). The teachers in these grades participated in the Focus intervention in the academic year 2011–2012. Table 3 presents an overview of the number of students and their background

Table 3
Descriptive statistics of the samples for the grade approach.

	Grade 2 2010–2011		Grade 3 2010–2011		Grade 4 2011–2012		Grade 5 2011–2012	
	Experimental n (%)	Control n (%)	Experimental n (%)	Control n (%)	Experimental n (%)	Control n (%)	Experimental n (%)	Control n (%)
Gender								
Boys	533 (49)	397 (49)	515 (48)	419 (51)	480 (46)	431 (53)	516 (52)	411 (48)
Girls	547 (51)	412 (51)	554 (52)	395 (49)	554 (54)	386 (47)	473 (48)	437 (52)
Socio-economic status								
0.0	825 (76)	535 (66)	802 (75)	558 (69)	775 (75)	548 (67)	753 (76)	570 (67)
0.3	102 (9)	58 (7)	106 (10)	75 (9)	98 (9)	75 (9)	94 (10)	65 (8)
1.2	53 (5)	41 (5)	68 (6)	26 (3)	75 (7)	26 (3)	60 (6)	34 (4)
Unknown	100 (9)	175 (22)	93 (9)	155 (19)	86 (8)	168 (21)	82 (8)	179 (21)
Total	1080	809	1069	814	1034	817	989	848

characteristics in grades two and three in the academic year 2010–2011, and in grades four and five in the academic year 2011–2012.

Multilevel analyses were conducted for the analysis of the Focus effects per *grade*. Two hierarchical levels were included in the models: students (level 1) and schools (level 2). In these models, students' scores at the post-test were used as the dependent variable. In Model 1, the experimental condition, pre-test scores, and student background characteristics were used as covariates. Interactions between the experimental condition, the pre-test, and student background variables were also included in the model. This model assessed the main effect of the Focus intervention and its differential effects for particular subgroups. In Model 2, school composition in terms of the percentage low-SES students per school was added to the analyses. This model tested whether the effects of the Focus-intervention were dependent of the schools' composition. The effects of the Focus intervention were tested by means of one-sided significance tests using $\alpha = 0.05$. For the sake of comparison effect sizes were computed for the main effects from the multilevel models as the experimental-control difference in adjusted posttest scores divided by the student-level pooled standard deviation (Hedges, 2007; ; van Kuijk, 2014).

In case of missing values in the predictor variables, students were excluded from these analyses. The percentage of excluded students ranged between 16.24% (294 students in grade 4 2011–2012) and 19.01% (362 students in grade 2 2010–2011). Exclusion of students was mainly due to missing information on their socio-economic background (13.7%–15.3%) and, to a smaller extent, by missing pretest information (3.4%–6.2%). The patterns of missing values differed for socio-economic background and pretest data. Pattern differences in missing values for these two predictor variables are explained based on the grade 5 2011–2012 data, but the pattern is similar for the other grades. At the school level complete information on socio-economic background is available for 75.9% of the schools; in case of incomplete information the most frequent pattern is missing information for all students (16.5%). For the pretest complete information on the school level is available for 63.3% of all schools, in the case of incomplete school level information it was usually only one student for whom the information is missing (22.8%).

For all four grades students excluded from the analyses were compared to students included from the analyses on the posttest scores. These tests revealed no significant differences between included and excluded students; grade 2 2010–2011 $t(1902) = 0.880$, $p = 0.379$, $d = 0.040$; grade 3 2010–2011 $t(1849) = 1.810$, $p = 0.070$, $d = 0.080$; grade 4 2011–2012 $t(1808) = -1.219$, $p = 0.223$, $d = 0.057$; grade 5 2011–2012 $t(1821) = -0.158$, $p = 0.875$, $d = 0.007$. In all cases the differences between included and excluded students were insignificant and very small. We therefore did not expect an impact of attrition on our results.

Table 4
Multi-level growth curve models for the first and second student cohort.

	Cohort 1						Cohort 2					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	<i>B</i>	<i>SEb</i>	<i>b</i>	<i>SEb</i>	<i>b</i>	<i>SEb</i>	<i>B</i>	<i>SEb</i>	<i>b</i>	<i>SEb</i>	<i>b</i>	<i>SEb</i>
Fixed Part												
Intercept	64.04 [*]	0.71	62.77 [*]	0.61	62.57 [*]	0.92	80.58	0.62	79.95 [*]	0.51	79.89 [*]	0.80
Time ¹			1.43 [*]	0.01	1.43 [*]	0.10			1.26 [*]	0.00	1.04 [*]	0.08
Time ²			−0.01 [*]	0.00	−0.01	0.01			−0.01 [*]	0.00	−0.02 [*]	0.00
Condition (experimental)					−0.65	1.27					0.11	1.10
Start intervention*Time ¹					0.09	0.16					0.05	0.14
Start intervention*Time ²					−0.01	0.00					0.02	0.03
Condition*Time ¹					−0.06	0.13					−0.11	0.11
Condition*Time ²					−0.00	0.01					−0.01	0.01
Start intervention*Condition*Time ¹					0.23	0.22					0.03	0.18
Start intervention*Condition*Time ²					−0.00	0.01					0.01 [*]	0.00
Random part												
School level	29.64	6.26	20.48	4.51	20.50	4.51	21.95	4.78	13.09	3.19	13.16	3.20
Student level	120.42	6.02	143.90	4.96	144.04	4.96	114.04	5.34	138.97	4.73	139.08	4.73
Measurement level	444.22	5.63	57.44	0.73	57.27	0.73	356.89	4.43	42.99	0.53	42.26	0.53
Model fit												
−2*loglikelihood	130,978.04		105,131.94		105,096.82		132,709.79		104,982.23		104,758.94	
Number of schools	82		82		82		82		82		82	
Number of students	1984		1984		1984		1976		1976		1976	
Number of measurements	14,410		14,410		14,410		14,941		14,941		14,941	

* $p < 0.05$.

3. Results

3.1. The cohort approach

Table 4 shows the results of the multilevel analyses for cohort 1 and cohort 2. The results of model 2 for both cohorts show a positive development in mathematics achievement, which somewhat decreases during the four years in which the students were studied. In Model 3, the effects of the Focus intervention were investigated. The results are presented in Fig. 4. The growth curve models for cohort 1 and cohort 2 differed to some extent, especially for the development in mathematics achievement during the Focus intervention. Growth for cohort 1 seems slightly stronger in the experimental relative to the control group at the start of the intervention, $b = 0.233$, $t(14399) = 1.06$, $p = 0.14$. In cohort 2, growth seems somewhat stronger at the end of the intervention for the experimental schools compared to the control schools, $b = 0.013$, $t(14930) = 3.25$, $p < 0.001$. This may be explained by the fact that for Cohort 2 the teachers only participated in the second intervention year. However, the results of the full analyses indicate that the differences between the experimental and

control groups, both before and during the intervention, are not statistically significant.

3.2. The grade approach

Next, the data per grade were analyzed using the last measurement before the start of the intervention as the pre-test. The results for students who were in grade two and three during the academic year 2010–2011 are presented in Table 5. Furthermore, the results of the analyses for the students in grade four and five during the academic year 2011–2012 are presented in Table 6.

For the students in grade two during the academic year 2010–2011, we found the following results with respect to the effects of the Focus intervention. No significant main effect of the Focus intervention was found, $b = -0.033$, $t(1533) = -0.03$, $p = 0.51$, $ES < 0.01$ (Model 1). Furthermore, an effect was found for the two-way interaction between the experimental condition and students' pre-test scores, $b = -0.080$, $t(1533) = -2.05$, $p = 0.02$ (Model 1). The negative interaction term implies that the effects of the intervention were larger

Fig. 4. Growth curves for the first and the second student cohorts.

Table 5
Multilevel models for grade 2 and grade 3 in academic year 2010–2011.

	Grade 2				Grade 3			
	Model 1		Model 2		Model 1		Model 2	
	<i>B</i>	<i>SEb</i>	<i>b</i>	<i>SEb</i>	<i>B</i>	<i>SEb</i>	<i>b</i>	<i>SEb</i>
Fixed Part								
Intercept	63.517 [*]	0.839	63.421 [*]	0.840	78.525 [*]	0.679	78.561 [*]	0.683
Condition (experimental)	−0.033	1.113	0.220	1.126	1.395	0.915	1.326	0.929
Pre-test	0.772 [*]	0.030	0.773 [*]	0.030	0.795 [*]	0.026	0.794 [*]	0.026
SES (0.3)	−4.576 [*]	1.584	−4.779 [*]	1.602	−2.380 [*]	1.116	−2.286	1.138
SES (1.2)	−4.370 [*]	1.964	−4.773 [*]	2.016	−6.529 [*]	1.899	−6.410 [*]	1.920
Condition*Pre-test	−0.080 [*]	0.039	−0.081 [*]	0.039	−0.033	0.033	−0.032	0.033
Condition* SES (0.3)	2.017	1.969	2.426	1.998	0.217	1.505	0.072	1.540
Condition* SES (1.2)	−0.507	2.558	0.143	2.613	4.235 [*]	2.197	4.080 [*]	2.220
Percentage SES (0.3 and 1.2)			0.040	0.047			−0.016	0.038
Condition*Percentage SES			−0.076	0.063			0.024	0.052
Random part								
School	12.431	3.167	12.173	3.120	8.680	2.167	8.633	2.167
Student	104.805	3.855	104.767	3.854	67.055	2.470	67.055	2.470
Model fit								
−2*log likelihood:	11,629.463		11,627.999		10,915.755		10,915.526	
Number of schools	68		68		67		67	
Number of students	1542		1542		1538		1538	

* $p < 0.05$

for students with lower pre-test scores. The interactions between experimental condition and student weights were not statistically significant, the same applies to school composition effects.

For the students in grade three in the academic year 2010–2011, we found the following results for the effects of the Focus-intervention. No significant main effect of the Focus intervention was found, $b = 1.395$, $t(1529) = 1.52$, $p = 0.08$, $ES = 0.17$ (Model 1). The two-way interaction between experimental condition and student weight 1.2, however, was significant, $b = 4.235$, $t(1529) = 1.93$, $p = 0.03$ (Model 1). The interaction is positive, hence, the Focus intervention was more effective for this group of students.

For the students in grade four in the academic year 2011–2012, we

found similar results. No significant main effect of the Focus intervention was found, $b = -1.031$, $t(1507) = 1.62$, $p = 0.95$, $ES = -0.16$ (Model 1). Furthermore, the interactions between the experimental condition and student background were not statistically significant; the same holds for school composition effects.

For the students in grade five in the academic year 2011–2012, the results of Model 1 indicate a significant positive main effect of the Focus-intervention ($b = 1.530$, $t(1507) = 1.89$, $p = 0.03$, $ES = 0.24$) while none of the interactions were significant. These results changed in Model 2 when school composition was added to the model. In Model 2, the main effect was no longer significant ($b = 1.162$, $t(1505) = 1.47$, $p = 0.07$, $ES = 0.18$), however, the interaction with school

Table 6
Multilevel models for grade 4 and grade 5 in academic year 2011–2012.

	Grade 2				Grade 3			
	Model 1		Model 2		Model 1		Model 2	
	<i>B</i>	<i>SEb</i>	<i>B</i>	<i>SEb</i>	<i>B</i>	<i>SEb</i>	<i>b</i>	<i>SEb</i>
Fixed Part								
Intercept	91.818 [*]	0.474	91.885 [*]	0.467	104.300 [*]	0.612	104.335 [*]	0.589
Condition (experimental)	−1.031	0.634	−1.077	0.627	1.530 [*]	0.809	1.162	0.790
Pre-test	0.741 [*]	0.020	0.740 [*]	0.020	0.939 [*]	0.022	0.938 [*]	0.022
SES (0.3)	−2.415 [*]	0.867	−2.126 [*]	0.887	−1.975 [*]	0.912	−1.963 [*]	0.922
SES (1.2)	−3.772 [*]	1.585	−3.438 [*]	1.601	−2.177 [*]	1.292	−2.110 [*]	1.308
Condition*Pre-test	−0.012	0.026	−0.011	0.026	0.002	0.029	0.004	0.029
Condition* SES (0.3)	1.378	1.173	1.142	1.206	0.268	1.191	−0.058	1.205
Condition* SES (1.2)	1.754	1.799	1.453	1.819	2.382	1.616	2.214	1.629
Percentage SES (0.3 and 1.2)			−0.040	0.026			−0.007	0.033
Condition*Percentage SES			0.034	0.035			0.075 [*]	0.044
Random part								
School	3.446	0.988	3.217	0.945	7.765	1.779	6.846	1.602
Student	40.804	1.513	40.812	1.513	40.186	1.499	40.227	1.501
Model fit								
−2*log likelihood:	9988.922		9986.569		9945.601		9940.724	
Number of schools	72		72		81		81	
Number of students	1516		1516		1507		1507	

* $p < 0.05$

composition was significant, $b = 0.075$, $t(1505) = 1.62$, $p = 0.04$. The interaction indicates that the effect of the Focus-intervention is dependent on the schools' composition. Larger effects were found in schools with relatively more low-SES students.

4. Discussion

Our results show that schools exposed to the Focus intervention did not benefit much from it in terms of their students' performance in mathematics. No main intervention effects were found; in all cases effect sizes for the main effects were small. However, three analyses yielded statistically significant interaction effects, indicating that the effects were more positive for students with lower pre-test scores and students with a lower socio-economic status. This finding is similar to the findings from other DBDM interventions (Keuning & van Geel, 2016). An explanation for the stronger intervention effects on low-SES students and on low-SES students may be that teachers, based on the performance data of their students, are inclined to focus their actions on those students who in their view most need support. As such it could be that high and average performing students do not benefit from improved teacher DBDM skills. These findings are contrary to our expectations, as data-based decision making is supposed to enable greater provision of tailor-made instruction as the data can show to teachers how much each student progresses, and which elements of subject matter content are not yet mastered by him/her.

Why did not students show more progress in the experimental group of schools? It is unlikely that the lack of a DBDM effect was caused by differences in prior achievement levels as the student performance levels at the start of the intervention were similar in the experimental and control group as can be seen in the multilevel growth curve models. Also propensity score matching was – although based on available school background data – applied in order to select appropriate control schools. The type of school inspection a school receives reflects the school's quality in a global way and does not reflect the quality of its individual teachers well. We cannot say with certainty that the quality of the teachers in the experimental and control group was sufficiently similar. In future studies full randomisation should avoid this problem.

The fact that we did not find the expected effects may have been caused by particular features of the Focus intervention. Two groups of teachers were trained separately, meaning that each group was trained for one year. Although we tried to support schools on multiple DBDM aspects, the Focus-intervention may have been too short to have an impact on teaching and student achievement. It is not easy to say how long teacher professional development trajectories intervention should be (Garet, Porter, Desimone, Birman, & Yoon, 2001; Van Veen et al., 2010) but we do have examples of similar DBDM interventions that lasted two years (for whole school teams) and that proved to be effective (Keuning & van Geel, 2016). Maybe the fact that the school team as a whole (including school leaders) did not participate at the same time has led to fragmentation in the implementation of DBDM.

However, the content of the intervention is probably the most important cause for the absence of the intended effects. DBDM presupposes that teachers master complex teaching skills (Inspectorate of Education, 2014; Van der Scheer, 2016). Such complex skills include the systematic analysis of the performance of all students and the differentiation of instruction in line with performance differences between students. In another publication, we reported that data literacy in the restricted sense (data analysis and data interpretation) is not the biggest problem (Staman, Visscher, & Luijten, 2014). School staff can master these data-analysis and data-interpretation skills if they are trained systematically. In this study we assumed that if we would teach

teachers how to analyse and interpret their student progress data correctly, and if we would assist them in setting goals and planning instruction in line with those goals, they would differentiate their classroom teaching in line with what they had found out about their students' progress. During the course of the intervention, we learned, however, that teachers experienced problems translating data analysis results into teaching strategies that provide students with more tailor-made instruction. It is indeed a challenge to deliver differentiated instruction in a class of 20 students who all have their specific characteristics and varying needs. We have supported teachers to help them connect the data with instructional plans (which included performance goals for all students, and instructional strategies for accomplishing these goals). These instructional plans were designed on paper by teachers and we provided every teacher with feedback on the quality of the plans twice an academic year. However, generating a plan does not ensure that the plan is executed well in the classroom. With about 700 teachers participating in our intervention, we could not coach teachers in the classroom in implementing their instructional plans.

Professional development trajectories have their limits in terms of the time teachers can spend on them, and in terms of their costs (the costs of involving coaches, and of substituting participating teachers). At the same time, digital feedback to teachers is becoming more and more important. In our intervention, teachers only received feedback on how well their students were performing twice a year but digital systems nowadays provide teachers with real-time feedback on student progress. This enables differentiated instruction but makes the challenge even bigger: continuous feedback asking for continuous differentiation.

Dutch teachers receive little training in differentiating their instruction during their pre-service training and almost no support for differentiating instruction in response to feedback from digital systems used in the classroom. It looks impossible to equip teachers for this in in-service training trajectories. Modern pre-service teacher training should include the utilization of (digital) student data, especially the deliberate answering of 'what-if questions': how to follow-up with instruction if that students with a specific profile perform in a specific way? If that is done effectively, in-service teacher training can optimize teachers' instruction by building on student data. As a result of this finding, we are now running a cognitive task analysis, investigating what it takes to do differentiation in the classroom well and, based on that knowledge, we are developing a teacher-training course for acquiring those competencies in a systematic way. The task analysis and training trajectory is based on the Four-Component Instructional Design Model for Complex Cognitive Skills (Van Merriënboer & Kirschner, 2013), an approach for analysing and acquiring complex skills that has been validated in many other areas (e.g. the training of medical specialists, software engineers, air traffic controllers, etc.).

Another way of promoting the intended effects of DBDM could be the further development of DBDM at the school level. DBDM effects within schools can be strengthened if not only individual teachers but also whole school teams are supported at the same time and encouraged to use data for evaluating the results in the various grades on a regular basis, and for finding ways in which school performance can be improved (the learning organization).

This study has deepened our insight into what it takes to work in a DBDM manner. The decomposition of DBDM into four components showed that the final step of implementing instructional plans in the classroom is the most challenging one, and that we cannot expect the average teacher to be able to do this without being trained for this purpose. This knowledge is important if we want teachers and students to benefit as much as possible from their feedback systems.

Appendix A. The topics and activities per meeting for the intervention years 2010–2011 and 2011–2012.

Academic year 2010–2011	Academic year 2011–2012	Academic year 2011–2012 (also school leader and academic coach)
Introductory meeting <ul style="list-style-type: none"> - Overview Focus project 	Meeting 1: Self-evaluation and how to teach mathematics <ul style="list-style-type: none"> - Brief overview of the Focus project - How to use the student monitoring system (example using a fictitious school) - Analyses student progress data with student monitoring system - Math ability: How to follow up after the test? - Math ability of students in the upper grades - Evaluation of and reflection on meeting 1 	Meeting 1: Results of the end-of-year test & optimizing DBDM <ul style="list-style-type: none"> - Starting situation for academic year 2011–2012 - Developing a teacher coaching plan - Evaluation of and reflection on meeting 1
Meeting 1: Self-evaluation <ul style="list-style-type: none"> - How to use the student monitoring system (example using a fictitious school) - Analyses student progress data with student monitoring system - Evaluation of and reflection on meeting 1 	Meeting 2: Developing an action plan <ul style="list-style-type: none"> - Reflection on the homework made - Goal setting theory - Features of strong math schools - Developing an action plan ('group plan') - Evaluation of and reflection on meeting 2 	Meeting 2: The support plan <ul style="list-style-type: none"> - Presentation of the teacher coaching plan - Exchange experiences group 2010–2011 - Evaluation of and reflection on meeting 2
Meeting 2: How to teach mathematics <ul style="list-style-type: none"> - Reflection on the homework made - Math ability: How to follow up after the test? - Example of a learning progression: Number sense - Evaluation of and reflection on meeting 2 	Meeting 3: The group plan <ul style="list-style-type: none"> - Reflection on the group plan - Optimizing the group plan - Mechanisms for performance improvement - Evaluation of and reflection on meeting 3 	Meeting 3: Optimizing DBDM <ul style="list-style-type: none"> - Presentation results Mid-year test academic year 2011–2012 - Evaluation of the support plan - Adapting the support plan - Evaluation of and reflection on meeting 3
Meeting 3: Goal setting <ul style="list-style-type: none"> - Reflection on the homework made - Goal setting theory - Features strong math schools - Developing an action plan ('group plan') - Evaluation of and reflection on meeting 3 	Meeting 4: The effects of DBDM <ul style="list-style-type: none"> - Reflection on the extent to which the performance goals had been accomplished - Evaluation of the degree of DBDM in the school - Adapting the group plan - Evaluation of and reflection on meeting 4 	Meeting 4 (= meeting 6 for teachers): The effects of two DBDM years <ul style="list-style-type: none"> - Reflection on the extent to which the performance goals had been accomplished - Evaluation of the degree of DBDM in the school - How to continue DBDM after Focus - Evaluation of and reflection on meeting 4 and the whole Focus-project
Meeting 4: The group plan <ul style="list-style-type: none"> - Reflection on the group plan - Optimizing the group plan - Mechanisms for performance improvement - Evaluation of and reflection on meeting 4 	Meeting 5: Optimizing of DBDM <ul style="list-style-type: none"> - Optimizing the adapted group plan - Optimizing DBDM in the school and the classroom - Evaluation of and reflection on meeting 5 	
Meeting 5: The first DBDM results <ul style="list-style-type: none"> - Reflection on the extent to which the performance goals had been accomplished - Evaluation of the degree of DBDM in the school - Adapting the group plan - Evaluation of and reflection on meeting 5 	Meeting 6: The effects of two DBDM years <ul style="list-style-type: none"> - Reflection on the extent to which the performance goals had been accomplished - Evaluation of the degree of DBDM - How to continue DBDM after the Focus project - Evaluation of and reflection on meeting 6 and the whole Focus project 	
Meeting 6: Optimizing DBDM <ul style="list-style-type: none"> - Optimizing the adapted group plan - Optimizing DBDM in school and classroom - Evaluation of and reflection on meeting 6 		
Meeting 7: The effects of 1 DBDM year <ul style="list-style-type: none"> - Reflection on the extent to which the performance goals had been accomplished - Evaluation of the degree of DBDM in the school - How to continue DBDM after the Focus project - Evaluation of and reflection on meeting 7 and the whole Focus project 		

References

- Carlson, D., Borman, G. D., & Robinson, M. (2011). A multi-state district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, 33, 378–398. <http://dx.doi.org/10.1080/00220973.2012.699899>.
- Dunn, K. E., Airola, D. T., Lo, W. J., & Garrison, M. (2013). Becoming data driven: The influence of teachers' sense of efficacy on concerns related to data-driven decision making. *The Journal of Experimental Education*, 81(2), 222–241. <http://dx.doi.org/10.1080/00220973.2012.699899>.
- Faber, J. M., & Visscher, A. J. (2014). *Digitale leerlingvolgsystemen. Een review van de effecten op de leerprestatie [Digital Student monitoring systems: A review of the effects on student performance]*. Enschede: Universiteit Twente (commissioned by Kennisnet). Retrieved, April 25, 2014, from: <http://www.kennisnet.nl/onderzoek/nieuws/digitale-leerlingvolgsystemen-een-review-van-de-effecten-opleerprestaties>.
- Fullan, M. (2007). *The new meaning of educational change*. New York and London: Teacher College Press.
- Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945. <http://dx.doi.org/10.3102/00028312038004915>.
- Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making (NCEE 2009-4067)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341–370.
- Hop, M., & Engelen, R. (2017). *Wetenschappelijke verantwoording rekenen-Wiskunde 3.0 voor groep 6*. Arnhem, The Netherlands: Cito.
- Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. In P. A. Moss (Vol. Ed.), *Evidence and decision making*. Vol. 106, (pp. 105–131). Malden, MA: Blackwell Publishing Inc.
- Inspectorate of Education (2009). *De staat van het onderwijs: Onderwijsverslag 2007/2008 [The State of education: Educational report 2007/2008]* De staat van het onderwijs: Onderwijsverslag 2007/2008 [The state of education: Educational report 2007/2008] Utrecht: Author.
- Inspectorate of Education (2013). *De staat van het onderwijs: Onderwijsverslag 2011/2012. [The state of education: Educational report 2011/2012]* Utrecht: Author.
- Inspectorate of Education (2014). *De staat van het onderwijs: Onderwijsverslag 2012/2013. [The state of education: Educational report 2012/2013]* Utrecht: Author.
- Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Scientific justification of the LOVS mathematics tests for Grade 1 to 6*. Arnhem, Netherlands: Cito.
- Keuning, T., & Van Geel, M. J. M. (2012). Focus projects II and III. The effects of a training in 'achievement oriented work' for primary school teams. *Poster presented at the International ICO fall school*.
- Keuning, T., & van Geel, M. J. M. (2016). *Implementation and effects of a schoolwide data-based decision making intervention: A large-scale study*. Enschede: Universiteit Twente.
- Data-based decision making: An overview. In M. K. Lai, & K. Schildkamp (Eds.), *Data-based decision making in education*. Dordrecht: Springer.
- Ledoux, G., Blok, H., Boogaard, M., & Krüger, M. (2009). *Opbrengstgericht werken; over de waarde van meetgestuurd onderwijs [Data-based decision making: on the value of measurement based education]*. Amsterdam: SCO-Kohnstamm Instituut.
- Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30–37. <http://dx.doi.org/10.3102/0013189X12459803>.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/occasional_papers/2006/RAND_OP170.pdf.
- Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, 114, 30–48.
- McCaffrey, D. F., & Hamilton, L. S. (2007). *Value-added assessment in practice: Lessons from the pennsylvania value-added assessment system pilot program*, Vol. 506. Rand Corporation.
- Ministry of Education (2007). *Scholen voor Morgen [Schools for tomorrow]*. Den Haag: Author.
- Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2009). *A users guide to MLwiN version 2.10*. Centre for Multilevel Modelling, University of Bristol.
- Ritzema, L. (2015). *Professional development in data use: The effects of primary school teacher training on teaching practices and students' mathematical proficiency*. Groningen: GION.
- 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.
- Slavin, R. E., Cheung, A., Holmes, G. C., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal*, 50, 371–396.
- Staman, L., Visscher, A. J., & Luyten, H. (2014). The effects of professional development on the attitudes, knowledge and skills for data-driven decision-making. *Studies in Educational Evaluation*, 42, 79–90. <http://dx.doi.org/10.1016/j.stueduc.2013.11.002>.
- Supovitz, J. A., & Turner, H. M. (2000). The effect of professional development on teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37, 963–980.
- Timperley, H. (2008). *Teacher professional learning and development*. Brussels: International Academy of Education/International Bureau of Education.
- Van Kuijk, M. F. (2014). *Raising the bar for reading comprehension*. Groningen: GION.
- Van Merriënboer, J. J. G., & Kirschner, P. (2013). *Ten steps to complex learning. A systematic approach to four-component instructional design*. Routledge: New York, London.
- Van der Scheer, E. A. (2016). *Data-based decision making put to the test*. Universiteit Twente; Enschede.
- Van Veen, K., Zwart, R., Meirink, J., & Verloop, N. (2010). *Professionele ontwikkeling van leraren: een reviewstudie naar effectieve kenmerken van professionaliseringsinterventies van leraren*. ICLON/Expertisecentrum Leren van Docenten; Leiden, Reviewstudie in opdracht van en gesubsidieerd door NWO-PROO (Grant no. 441-080353).
- Visscher, A., Peters, M., & Staman, L. (2010). *Het Focus-project: Opbrengstgericht werken op basis van prestatiefeedback [The Focus project: Data-based decision making based on performance feedback]*. Reken-wiskundeonderwijs: Onderzoek, ontwikkeling, praktijk, Vol. 29, 55–60.
- Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement*, 19, 239–259. <http://dx.doi.org/10.1080/09243450802246376>.