Abstract Title Page

Title:

The Effects of a Schoolwide Data-Based Decision Making Intervention on Elementary Schools' Student Achievement Growth for Mathematics and Spelling

Authors and Affiliations:

Trynke Keuning
University of Twente, The Netherlands

Marieke van Geel
University of Twente, The Netherlands

Adrie Visscher
University of Twente, The Netherlands

Jean-Paul Fox
University of Twente, The Netherlands

Abstract Body

Background

Around the world, during the last decade policy makers increasingly emphasize the use of data in education to enhance student achievement (Orland, 2015; Schildkamp, Ehren, & Lai, 2012). As a result, the number of reform initiatives to promote 'data-based decision making' (DBDM) or 'data-driven decision making' (DDDM) has increased rapidly (e.g. Boudett, City, & Murnane, 2005; Carlson, Borman, & Robinson, 2011; Love, Stiles, Mundry, & DiRanna, 2008; Schildkamp, Poortman, & Handelzalts, 2015; Slavin, Cheung, Holmes, Madden, & Chamberlain, 2012). The idea of using student achievement data for evaluating student progress, for providing tailor-made instruction, and for developing strategies for maximizing performance in order to positively influence student outcomes, seems straightforward. However, evidence on the effectiveness of DBDM reform is scarce. In large-scale studies the effect of data-use interventions on student achievement so far were insignificant (Henderson, Petrosiono, Guckenburg, & Hamilton, 2007; Quint, Sepanik, & Smith, 2008), or small (Carlson et al., 2011; Konstantopoulos, Miller, & van der Ploeg, 2013; May & Robinson, 2007). This does not necessarily imply that data-use in education is not effective, but rather suggests that more research is needed on how data-use can reach its full potential (Kaufman, Graham, Picciano, Popham, & Wiley, 2014).

Next to the well-known features of effective teacher professional development such as collective participation, a clear link between the intervention content and educational practice, enough time to practice newly learned methods (Desimone, 2009; Timperley, 2008; Van Veen, Zwart, & Meirink, 2011), two matters are specifically important in developing a DBDM intervention.

First, DBDM interventions should include all DBDM components, in a coherent and consistent way. As Kaufman (2014) states: "While identifying and analyzing data lays the groundwork for impactful improvements to student learning, the resulting actions and progress monitoring will ultimately determine the efficacy of DDDM efforts" (p. 341). In Figure 1, DBDM is decomposed into four components (Keuning, van Geel, Visscher, Fox, & Moolenaar, in press). The first component, analyzing and evaluating data, is only meaningful when it is a part of the entire DBDM cycle. Based on the insights gained from the analysis of data, SMART and challenging goals should be set. Next, strategies are chosen to accomplish these goals, and finally the chosen strategy should be executed. Since DBDM ideally is carried out as a systematic approach, data is also supposed to be used for monitoring and evaluating the effects and outcomes of the implemented strategy, for evaluating the extent to which goals have been achieved, and for making new data-informed decisions. As all components are related to each other, in order for DBDM interventions to be meaningful and effective, these interventions should include *all* DBDM components.

Second, interventions should take both the school level as well as the teacher level into account. Researchers found that despite of schools and/or districts actively promoting DBDM, teachers felt unprepared to work with data. Even when they learned how to analyze and interpret data, they did not change their classroom practice (Means, Padilla, & Gallagher, 2010; Schildkamp & Kuiper, 2010). An explanation might be that DBDM initiatives until now did not affect teachers much, and therefore showed only minor effects on classroom practice, whereas teachers can make a difference at the classroom level (Borko, 2004). According to Kaufman et al. (2014) there is a need for research on "how to improve and even speed up adoption of effective data use practices in school settings" (p. 343).

(please insert Figure 1 here)

At the University of Twente in the Netherlands, a DBDM intervention was developed in which whole school teams participate in the training. DBDM was introduced as a systematic approach, teachers learned how to analyze data, to set goals and to choose instructional strategies based on these data, and next to alter their instruction in the classroom accordingly. In 2011 a first group of 53 elementary schools participated in this DBDM intervention and showed promising results (Van Geel, Keuning, Visscher, Fox, 2015). The analysis of student achievement data for mathematics revealed a significant student achievement gain of approximately one extra month of schooling during the two intervention years for all students involved. Furthermore, the results indicated that the intervention especially improved the performance of students in low-SES schools (Van Geel, Keuning, Visscher, Fox, 2015). In 2012 a new cohort of schools started the intervention, the study reported on in this abstract was aimed at evaluating the intervention effects of this new cohort of schools.

Purpose

As Borko (2004) stated, in order to provide high-quality professional development for all teachers, professional development programs should be evaluated in different settings and with different program facilitators. Therefore, for the current study a similar intervention as the intervention starting in 2011 was implemented and evaluated in a new cohort of 40 elementary schools . This study expands the previous study as student achievement for both mathematics and spelling was analyzed. As such, this study can be considered to be a conceptual replication study (Makel & Plucker, 2014; Schmidt, 2009) with the aim to generalize findings and to broaden our understanding of the effects of this DBDM intervention.

Setting

Data for this study were gathered from 40 elementary (K-6) schools in the Netherlands which participated in the DBDM intervention from August 2012 until July 2014. Student achievement data covering the period August 2010 until July 2014 were retrieved from schools' student monitoring systems.

Participants

Characteristics of the 40 participating schools are presented in Table 1. Schools were supposed to first choose one subject (mathematics, spelling, vocabulary, or reading) to focus on during the intervention. After one year, they could add another subject, or stick to the same subject. After one and a half year schools again could choose to work on a new subject, or not. This approach resulted in different intervention trajectories. Five schools which did not include mathematics into their trajectory were removed from the sample for the analysis of mathematics achievement. For spelling, 12 schools did not include spelling and therefore were removed from the analysis of the spelling results. Next, students of whom only the data from one measurement were available were removed from the sample. This resulted in a sample of 8,396 unique students for mathematics, and 6,615 unique students for spelling. Table 2 presents the characteristics of these students.

(please insert Table 1 & 2 here)

Intervention

The DBDM intervention was a two-year training course for entire Dutch elementary school teams (all teachers as well as the members of the management team such as the school leader and deputy director), aimed at implementing and sustaining DBDM in the whole school organization, by systematically following the DBDM cycle as shown in Figure 1.

The first year of the intervention included seven team meetings aimed at developing DBDM knowledge and skills. The first four meetings were primarily aimed at DBDM related knowledge and skills: analyzing and interpreting test score data from the student monitoring system, diagnosing learning needs, setting performance goals, and developing instructional plans. Before the fifth meeting teachers had executed the instructional plans in the classroom, and, based on students curriculum-based tests, classwork, homework and classroom observations, they had adjusted those plans if necessary. By the fifth meeting, the DBDM cycle had been completed for the first time, and student achievement data were then discussed in a team meeting. During this meeting teachers shared their effective and ineffective classroom practices. Meeting six focused on collaboration among team members by preparing them for observing each other's lessons; either to learn from the colleague they visited, or to provide him/her with feedback. In the last meeting of the school year, the DBDM cycle was completed for the second time as student results and classroom practices were evaluated again. Furthermore, teachers made an instructional plan for the next school year (and for the teacher(s) of that year), and also provided information about the class to the new teacher. In addition to the seven meetings, teachers were provided with feedback by the external trainer on both the way they had analyzed and interpreted data as well as on the quality of their instructional plans.
The second intervention year was aimed at deepening, sustaining and broadening DBDM within the school and included 5 meetings, in which new subjects were introduced (optional for schools). The DBDM cycle was completed again twice that year. Furthermore, two coaching sessions were included in this second school year, in which the DBDM trainer observed teachers' classroom instruction and provided them with feedback

Research Design

A multiple single-subject design was used to investigate the effect of this DBDM intervention on student achievement growth, and to investigate patterns in DBDM effectiveness based on background variables at both the school and the student level. Each school was measured repeatedly over time, before the intervention period (the control phase) and during the intervention period (the treatment phase). The purpose was to measure changes in scores (i.e., performance of each school), and to assess the impact of the intervention for each school. Jenson, Clarck, Kircher and Kristjansson (2007) and Van den Noortgate and Onghena (2003) advocate the use of hierarchical linear models to improve statistical inferences. The present research design extends the hierarchical linear modeling approach of single-subject design studies, by extending the level-1 model for the repeated measurements of a single-subject study. Through the joint modeling of multiple single-subject designs, each single-subject study of a school encompasses multivariate repeated measurements of students (representing the school), who are followed over time.

Data Collection and Analysis

Student performance on the standardized tests was scored on an ongoing ability scale per subject (math and spelling) for grade one to six (students aged six to twelve years old). For the two years before the intervention and the two intervention years, a maximum of eight measurements was observed out of the in total eleven measurements (two measurements per grade for grade years one to five, and one for grade six). The total number of observations for mathematics was 42,787; for spelling 35,361. An overview of test occasions is depicted in Figure 2. In addition to students' ability scores, at the student level data was collected on gender, SES category (high, medium, low), and the date of birth. Age was transformed based on the average age in months at the time of the test.

(please insert Figure 2 here)

Given the multilevel structure of the data, with measurements nested within students, and students nested within schools, the *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) in R (RCoreTeam, 2013) was used to perform linear mixed effects analyses, to investigate intervention effects on student achievement.

Growth was modeled by modeling heterogeneity in (average) student achievement, while accounting for differences between measurement occasions, and average test performance over students and schools. The differences in average achievement over grades were modeled as fixed effects, and student achievement and school achievement were allowed to vary across the general mean, by introducing student and school-specific random intercepts. Random effects were introduced for average achievement over grades three to five, and grades six to eight at the student level. At the school level, a random effect was introduced representing the variability in the effect of the intervention across schools. As mathematics and spelling are measured on different ability scales, the analyses for these two subjects were performed separately.

Findings

For both spelling and mathematics a significant intervention-effect was found. In Figure 3 the random intervention effect for each school was plotted against the random intercept. This Figure is based on the model which included all significant explanatory variables, but not the interaction-effects. Figure 3 for mathematics shows that the lower the school performed at the start of the intervention (reflected by a low intercept), the stronger the intervention-effect. This trend is less observable for spelling, as can be seen in Figure 4.

Including interaction-effects revealed that the positive intervention-effect for mathematics in particular yielded for students with low-SES, and high-SES (in comparison to medium-SES students). Additionally, schools with many low-SES students benefitted most from the intervention, compared to medium-SES and low-SES schools. For spelling no significant interaction effects were found.

Table 3 presents the results of the final models for both math achievement as well as spelling achievement.

(please insert Figure 3 & Table 3 here)

Conclusions

The current study contributes to the DBDM knowledge base by showing that a DBDM intervention in which whole school teams are actively involved, and in which all DBDM components are systematically executed can improve student outcomes. The study confirms the findings of the study by Van Geel et al. (2015) that mathematic outcomes improve especially for low-SES schools. Moreover, this study revealed that the DBDM intervention effects also hold for spelling. Interestingly, for spelling the effect of the intervention was not related to students' SES.

Appendices

Appendix A. References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [stat.CO]*.

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational Researcher*, *33*(8), 3–15. Retrieved from http://www.jstor.org/stable/3699979

Boudett, K. P., City, E., & Murnane, R. (2005). *Data Wise: A step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, MA: Harvard Education Press.

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378–398. doi:10.3102/0162373711412765

Desimone, L. M. (2009). Improving impact studies of teachers ' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Henderson, S., Petrosiono, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments af fect student achievement Measuring how benchmark assessments affect student achievement*. (Issues & Answers Report, REL 2007-No. 039). Washington, DC. Retrieved from http://ies.ed.gov/ncee/edlabs

Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-analyses, and single subject designs. *Psychology in the Schools*, *44*(5), 483–493. doi:10.1002/pits.20240

Kaufman, T., Graham, C., Picciano, A., Popham, J. A., & Wiley, D. (2014). Data-Driven Decision Making in the K-12 Classroom. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology SE - 27* (pp. 337–346). Springer New York. doi:10.1007/978-1-4614-3185-5_27

Keuning, T., van Geel, M., Visscher, A., Fox, J.-P., & Moolenaar, N. (n.d.). The Transformation of Schools' Social Networks During a Data-Based Decision Making Reform. *Teachers College Record*.

Konstantopoulos, S., Miller, S. R., & van der Ploeg, a. (2013). The Impact of Indiana's System of Interim Assessments on Mathematics and Reading Achievement. *Educational Evaluation and Policy Analysis*, *35*(4), 481–499. doi:10.3102/0162373713498930

Love, N., Stiles, K. E., Mundry, S., & DiRanna, K. (2008). *A data coach's guide to improve learning for all students: Unleashing the power of collaborative inquiry*. Thousand Oaks, CA: Corwin Press.

Makel, M. C., & Plucker, J. a. (2014). Facts Are More Important Than Novelty: Replication in the Education Sciences. *Educational Researcher*, *43*(6), 304–316. doi:10.3102/0013189X14545513

May, H., & Robinson, M. A. (2007). *A randomized evaluation of Ohio's personalized assessment reporting system (PARS)*. Madison, WI.

Means, B., Padilla, C., & Gallagher, L. (2010). *Use of education data at the local level: From accountability to instructional improvement*. Retrieved from http://www2.ed.gov/rschstat/eval/tech/use-of-education-data/use-of-education-data.pdf

Orland, M. (2015). Research and Policy Perspectives on Data Based Decision Making in Education. *Teachers College Recird*, *117*(4).

Quint, J. C., Sepanik, S., & Smith, J. K. (2008). *Using Student Data to Improve Teaching and Learning: Findings from an Evaluation of the Formative Assessments of Student Thinking in Reading (FAST-R) Program in Boston Elementary Schools* (p. 145). New York, NY: MDRC.

RCoreTeam. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Schildkamp, K., Ehren, M., & Lai, M. K. (2012). Editorial article for the special issue on data-based decision making around the world: from policy to practice to results. *School Effectiveness and School Improvement*, *23*(2), 123–131. doi:10.1080/09243453.2011.652122

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. doi:10.1016/j.tate.2009.06.007

Schildkamp, K., Poortman, C., & Handelzalts, A. (2015). Data teams for schoolimprovement. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*. doi:10.1080/09243453.2015.1056192

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. doi:10.1037/a0015108

Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2012). Effects of a Data-Driven District Reform Model on State Assessment Outcomes. *American Educational Research Journal*, *50*(2), 371–396. doi:10.3102/0002831212466909

Timperley, H. (2008). *Teacher professional learning and development*. International Academy of Education.

Van den Noortgate, W., & Onghena, P. (2003). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*, 325–346.

Van Geel, M., Keuning, T., Visscher, A., Fox, J.P. (2015) *Assessing the effects of a school wide data-based decision making intervention on student achievement growth in primary schools.* Manuscript submitted for publication.

Van Veen, K., Zwart, R., & Meirink, J. (2011). What makes teacher professional development effective? A literature review. In M. Kooy & K. Van Veen (Eds.), *Teacher learning that matters* (pp. 3–21). New York: Routledge.

Appendix B. Tables and Figures

Table 1. School Characteristics (N=40)

		N	(%)
School Size	Small (<150)	13	(32.5%)
(number of students)	Medium (150-350)	20	(50.0%)
	Large (>350)	7	(17.5%)
Urbanization	Rural	17	(42.5%)
	Suburban	16	(40.0%)
	Urban	7	(17.5%)
School SES	High	12	(30.0%)
	Medium	21	(52.5%)
	Low	7	(17.5%)
Main intervention subject	Math	21	(52.5%)
	Spelling	15	(37.5%)
	Reading	3	(7.5%)
	Vocabulary	1	(2.5%)
Trajectory Spelling	000	12	(30.0%)
Y1 – Y2 part 1 – Y2 part 2	001	2	(5.0%)
	011	11	(27.5%)
	100	12	(30%)
	110	1	(2.5%)
	111	2	(5.0%)
Trajectory Math	000	5	(12.5%)
Y1 – Y2 part 1 – Y2 part 2	001	1	(2.5%)
	010	1	(2.5%)
	011	12	(30.0%)
	100	11	(27.5%)
	110	3	(7.5%)
	111	7	(17.5%)

Table 2. Student Characteristics for Mathematics (N=8,396) and Spelling (N=6,615)

		Math:		Spelling	
		N	(%)	N	(%)
Gender	Boy	4214	(50.3%)	3333	(50.4%)
	Girl	4182	(49.8%)	3282	(49.6%)
Student SES	High (0.0)	6779	(80.7%)	5688	(86.0%)
	Medium (0.3)	688	(8.2%)	444	(6.7%)
	Low (1.2)	922	(11.0%)	476	(7.2%)
Number of observations per student	2	1740	(20.7%)	1165	(17.6%)
	3	676	(8.1%)	600	(9.1%)
	4	1508	(18.0%)	1095	(16.6%)
	5	659	(7.8%)	509	(7.7%)
	6	1194	(14.2%)	935	(14.1%)
	7	643	(7.7%)	639	(9.7%)
	8	1517	(18.1%)	1170	(17.7%)
	> 8	459	(5.5%)	502	(7.6%)

Table 3. Final Model For Mathematics And Spelling

	Mathematics:		Spelling	
	Est.	S.E.	Est.	S.E.
(Intercept)	34.26	2.12**	104.15	.42**
<u>Student level</u>				
Test end grade 3	11.91	.17**	6.40	.12**
Test mid grade 4	19.77	.19**	1.96	.12**
Test end grade 4	31.16	.19**	13.31	.12**
Test mid grade 5	39.33	.20**	18.14	.13**
Test end grade 5	47.10	.20**	22.24	.13**
Test mid grade 6	53.49	.23**	24.94	.15**
Test end grade 6	59.65	.23**	29.30	.15**
Test mid grade 7	67.81	.26**	3.82	.16**
Test end grade 7	72.75	.26**	32.28	.16**
Test mid grade 8	79.96	.30**	35.30	.19**
Intervention	.32	.49	.84	.18**
Student SES - high	6.45	.55**	3.12	.31**
Student SES - low	-.07	.69	.44	.42
Student gender (1=f)	-3.57	.29**	1.19	.15**
Student age (months)	.50	.02**	.07	.01**
Intervention*StudentSES high	.79	.32*		
Intervention * StudenSES low	1.05	.41*		
<u>School level</u>				
SchoolSize - large	.83	.87		
SchoolSize - small	-2.06	.73**		
Suburban	-2.98	.67**		
Urban	-3.93	.94**		
SchoolSESlow	-3.07	1.01**		
SchoolSEShigh	2.06	.95*		
TrajectRWRW010	-6.26	2.48*		
TrajectRWRW011	-8.56	1.97**		
TrajectRWRW100	-6.01	1.91**		
TrajectRWRW110	-7.08	2.03**		
TrajectRWRW111	-8.23	1.98**		
Intervention * SchoolSESlow	1.71	.76*		
Intervention * SchoolSEShigh	-.52	.68		
<u>Variance components</u>				
student level				
(Intercept)	174.37	13.21	31.64	5.62
Clust345	36.77	6.06	11.92	3.45
Clust678	71.70	8.47	25.18	5.02

school level				
(Intercept)	3.68	1.92	2.00	1.41
Intervention	2.59	1.61	.79	.89
residual	41.05	6.41	16.07	4.01

Note. Only variables with a significant effect were included in the final model.
Note. As mathematics and spelling are different constructs and measured on a different scale, effect sizes are not comparable

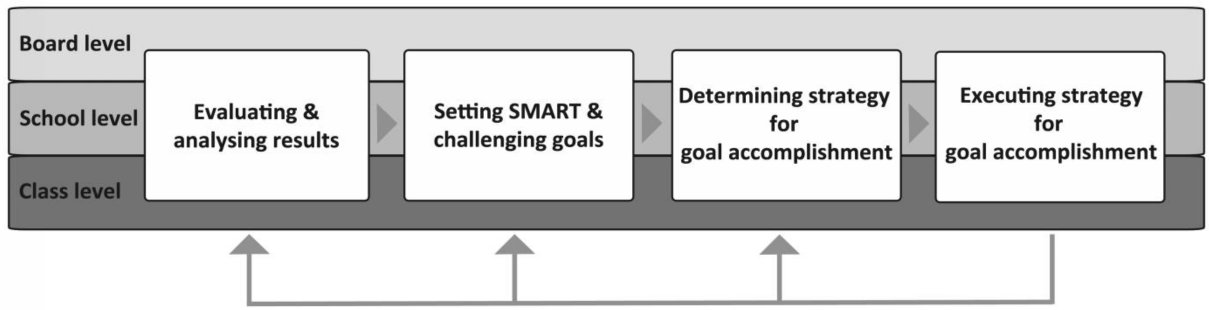

Figure 1. The DBDM cycle (Keuning et al., in press)

	Prior to intervention				During intervention			
	School Year 2010-2011		School Year 2011-2012		School Year 2013-2013		School Year 2013-2014	
	Mid (Febr)	End (June)	Mid (Febr)	End (June)	Mid (Febr)	End (June)	Mid (Febr)	End (June)
Grade 3	X	X	X	X	X	X	X	X
Grade 4	X	X	X	X	X	X	X	X
Grade 5	X	X	X	X	X	X	X	X
Grade 6	X	X	X	X	X	X	X	X
Grade 7	X	X	X	X	X	X	X	X
Grade 8	X	-	X	-	X	-	X	-

Figure 2. Overview of Measurement Occasions. Shadings Indicate Cohorts.

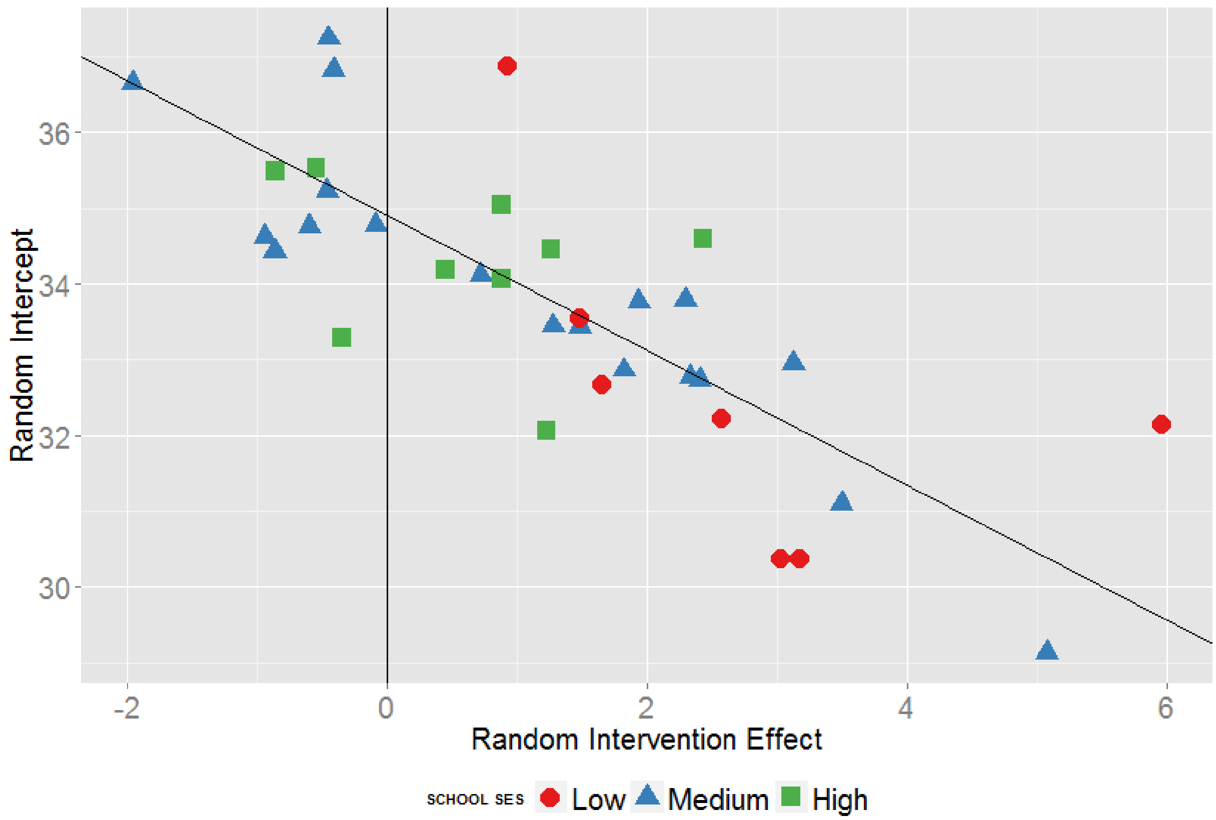

Figure 3. Random Intervention Effect Plotted Against Random Intercept for Mathematic Achievement.

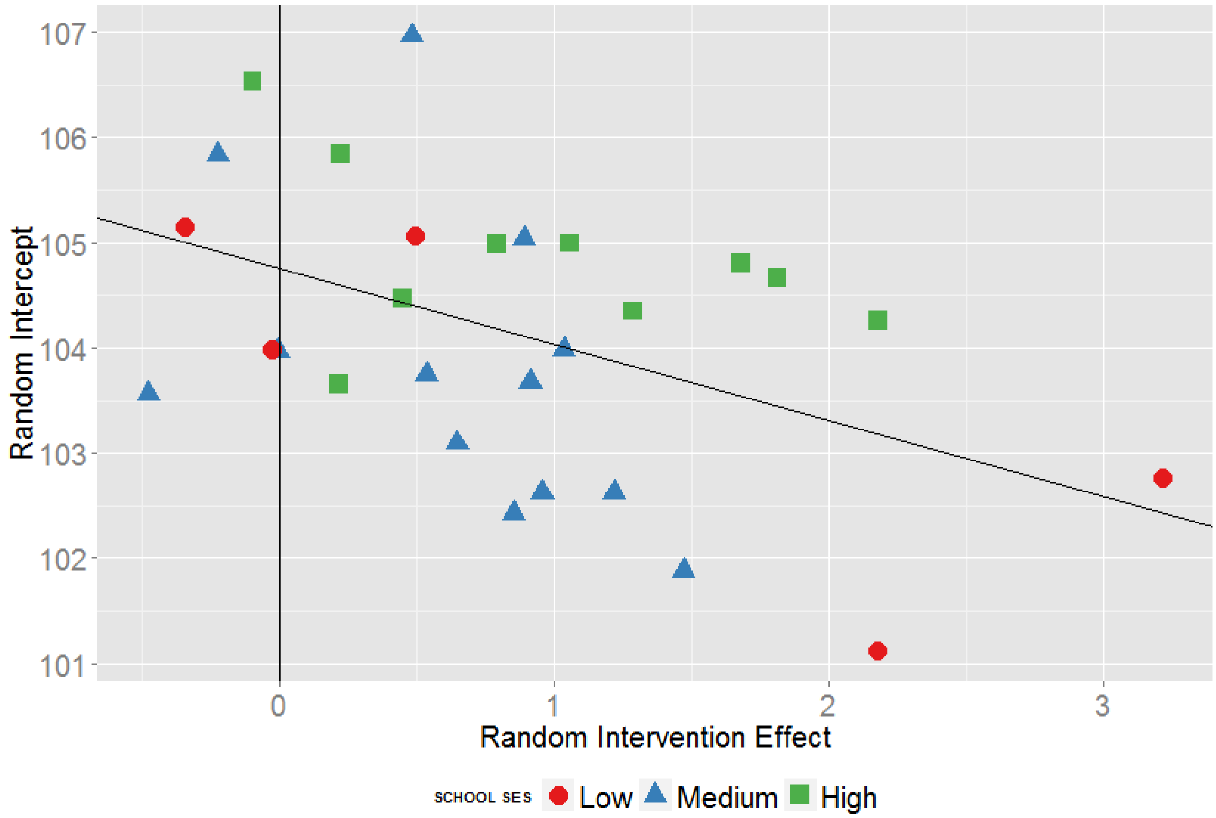

Figure 4. Random Intervention Effect Plotted Against Random Intercept for Spelling Achievement.