Abstract Title Page
Not included in page count.

Title: Building Measures of Instructional Differentiation from Teacher Checklists

Authors and Affiliations: Ryan Williams[1], Andrew Swanlund[1], Shazia Miller[1], Spyros Konstantopoulos[2], Arie van der Ploeg[1]

American Institutes for Research[1]
Michigan State University[2]

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Differentiated instruction (is commonly believed to be critical to improving the quality and efficiency of teachers' instructional repertoires (Fischer & Rose, 2001; Tomlinson, 2004). Tomlinson (2000) describes differentiation in four domains: content, process, product, and learning environment. Content differentiation involves varying instructional topics, for example, that students within a classroom would receive. Process differentiation involves teaching different students at different levels of difficulty. Product differentiation involves assigning different tasks to different students. Learning environment differentiation includes using different instructional groupings for students or clustering students based on ability.

Methods to measure instructional differentiation are not well refined. Some studies have relied on survey methods. Graham and colleagues (2008), for example, conducted a national survey of differentiated instruction in spelling. The authors constructed a set of survey items that asked teachers about the quantity and frequency of various instructional practices, including instructional adaptations for elementary school students of varying ability levels. Other studies have relied on classroom observations to examine differentiated instruction. For example, in a randomized experiment, VanTassel-Baska and colleagues (2008) used a structured classroom observation instrument to measure change in differentiated instruction over three years.

However, these methods have limitations (Rowan, Camburn, & Correnti 2004). Surveys often require teachers to recall instructional behavior over a long period of time (e.g. an entire academic year). This method is susceptible to retrospective self-report bias, especially when reporting on infrequent behaviors or events. Classroom observation systems avoid teacher self-reports entirely. But, trained observers cannot always observe the entirety of the instructional process or teachers' intents (e.g. which questions were assigned to which students or subtle adjustments to the ability levels of each student in the classroom). Rowan and Correnti (2009) also note that classroom observation systems will typically lack the generalizability of survey methods and that observations can be costly.

Another method for deriving measures of instructional differentiation is through the use of teacher checklists (or logs) (e.g. Rowan, Correnti, & Miller, 2002). Teacher checklists are commonly administered multiple times throughout an academic year, in hopes of capturing an instructional profile that spans the entire year. While checklists are more burdensome on teachers because of frequent administration, they are typically more generalizable than classroom observation methods and survey methods because of frequent administration. Checklists are less taxing on teacher memory because teachers focus on a specific date and complete the checklist on or around that date. Rowan and Correnti (2009) used teacher checklists to measure instructional practices in the Study of Instructional Improvement. In this study, teachers completed checklists at the end of specifically designated days, reflecting on instruction provided to eight randomly selected students in their classrooms.

Purpose / Objective / Research Question / Focus of Study:
Description of the focus of the research.

The purpose of this paper is to present a measure of instructional differentiation derived from teacher checklist data and apply it in the context of a randomized experiment that is part of a phased rollout of Diagnostic Assessment Tools statewide in Indiana.

Setting:
Description of the research location.

This work is part of the Indiana Diagnostic Assessment Tools randomized control trial (RCT), year 1. The experiment occurred in Indiana in 2009-2010 and included 50 randomly selected K-8 schools that had volunteered to be part of the intervention the previous spring. Only teachers or reading and mathematics in grades 2 and 5 were asked to complete the checklists.

Population / Participants / Subjects:
Description of the participants in the study: who, how many, key features, or characteristics.

This measurement study used data obtained from treatment and control school teachers of English language arts (ELA) and mathematics in grades 2 and 5. In grade 2 ELA, the sample included 42 control and 53 treatment teachers; in mathematics, 37 control and 42 treatment teachers. In grade 5 ELA, the sample was 45 control and 54 treatment teachers; in grade 5 mathematics, 41 control and 47 treatment teachers.

Intervention / Program / Practice:
Description of the intervention, program, or practice, including details of administration and duration.

Indiana rolled out two assessment systems. The first was Wireless Generation's *mCLASS:Reading 3D* and *mCLASS:Math* as the K–2 solution and the second was CTB/McGraw-Hill's *Acuity* product for Grades 3–8..

Research Design:
Description of the research design.

Teachers in control and treatment schools in grades 2 and 5 were asked to complete 16 instructional checklists throughout the 2009-2010 school year, roughly one every two weeks. Our staff developed four checklist versions, a grade 2 and 5 ELA checklist and a grade 2 and 5 mathematics checklist. The ELA checklists were based on Rowan and Correnti's checklist (2009).The mathematics checklists were developed by content experts, following the ELA model and guided by the Indiana mathematics standards. There were a total of 186 items for the grade 2 ELA checklist, 237 items for the grade 2 mathematics checklist, 186 items for grade the 5 ELA checklist, and 273 items for the grade 5 mathematics checklist. In each checklist, items were categorized by topic area. For example, the grade 5 math checklist had seven topic areas: number sense; computation; algebra and function; geometry; measurement; problem solving; and data analysis and probability. And, each topic area contained items related to teacher instruction, concepts and skills, student materials, and student activities. Teachers completed the checklists

online and results were stored on servers. Teachers were sent reminders when to complete the checklist, and follow ups if responses were not made in timely fashion.

Following the procedures described by Rowan and Correnti (2009), eight students were randomly selected by each teacher to focus on while completing the checklist. These same 8 students were used for the entire year.

For each student, for each checklist item, teachers indicated either if a student was instructed in a topic or not and whether they used a particular instructional practice or not. If they had taught particular content, they indicated whether they had taught that student at the remedial, regular, or enriched level.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Because of the complexities of the checklist data, our measure of differentiated instruction is an item response theory- (IRT) based measure. Each dichotomous item was coded as 1 if at least one student was taught something that other students were not, 0 otherwise. For polytomous items asking about the level of instruction each student was taught at (e.g. remedial, regular, enriched), each item was coded as 1 if students were taught at different levels, 0 otherwise. This measure makes use of all items pertaining to each of the mathematics and ELA topic areas (e.g. comprehension, number sense, etc.) items. Teacher (in time) item responses were fit to the one parameter Rasch model for binary responses to scale the response set, such that

$$P_{ni} = \frac{\exp(B_n - D_i)}{[1 + \exp\,(B_n - D_i)]},$$

where the estimated probability that a teacher *n* endorses item *i*, depends on *n*th teacher's estimated ability (level of differentiation) B_n, and the difficulty of the item D_i

For each grade and subject area, we analyzed the psychometric properties of this measure at the log-, teacher-, and school-level using the Rasch model and generalizability theory methods outlined by Rowan and Correnti (2009). Each reliability estimate indicates how well our measure separates units (i.e. teachers in time, teachers, and schools) along the latent dimension of differentiation.

Findings / Results:
Description of the main findings with specific details.

Our preliminary analyses indicate that the psychometric properties of our differentiation measures are generally strong. Log-, teacher-, and school-level reliability estimates are high (>.70) or moderately high (>.66) for each subject by grade grouping except grade 2 mathematics where the estimated log-level reliability estimate was .57 and the estimated school-level reliability estimate was .63.

When we applied this method to data from the RCT, we found notable differences between the treatment and control group. In general, both treatment and control groups rarely differentiated

their instruction. Interestingly, across all subjects and grades, the control group differentiated their instruction substantially more than the treatment group. The magnitude of these differences ranged from .11 standard deviations in grade 2 ELA to .44 standard deviations in grade 5 mathematics. In addition to computing overall estimates of differentiated instruction, we were able to compute longitudinal profiles of differentiated instruction that span the academic year for both groups. These profiles further illustrate the difference between the treatment and control group teachers.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

The results of this study indicate that teacher checklists can be used to build reliable measures of differentiated instruction. This approach may be generalized to look at differentiation in specific areas (e.g. within materials used to teach number sense). We will discuss additional applications of this method and ways it might be improved.

We are currently in the process of completing parallel analyses for year 2 of the RCT which also includes grade 6 ELA and mathematics. Substantive comparisons between the two sets of results will be made.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Fischer, K., & Rose, L. (2001). Webs of skill : How students learn. *Educational Leadership, 59*(3), 6-123.

Graham, S., Morphy, P., Harris, K. R., Fink-Chorzempa, B., Saddler, B., Moran, S., & Mason, L. (2008). Teaching spelling in the primary grades: A national survey of instructional practices and adaptations. *American Educational Research Journal*, *45*(3), 796 -825.

Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the Study of Instructional Improvement. *Educational Researcher*, *38*(2), 120.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *The Elementary School Journal*, 75–101.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schols. *The Teachers College Record*, *104*(8), 1525–1567.

Tomlinson, C. A. (2000). *Differentiation of instruction in the elementary grades*. Champaign, IL: ERIC Clearinghouse on Elementary and Early Childhood Education, University of Illinois.

Tomlinson, C. A. (2004). Research evidence for differentiation. *School Administrator,61*(7), 30.

VanTassel-Baska, J., Xuemei Feng, A., Brown, E., Bracken, B., Stambaugh, T., French, H., McGowan, S., et al. (2008). A Study of Differentiated Instructional Change Over 3 Years. *Gifted Child Quarterly*, *52*(4), 297 -312.

Appendix B. Tables and Figures

Not included in page count.