Abstract Title Page

Title: Making Sense of Unanticipated Results: Instructional Differentiation and the Indiana Diagnostic Assessment Study

Authors and Affiliations:

Ryan Williams (University of Memphis)
Andrew Swanlund (American Institutes for Research)
Shazia R. Miller (American Institutes for Research)
Spyros Konstantopoulos (Michigan State University)
Arie van der Ploeg (American Institutes for Research)
Abstract Body

Background / Context

The Indiana Diagnostic Assessment Study was a pair of two randomized control trials (RCTs) of the effects of a new benchmark assessment system. The logic behind these new tools is that they would provide teachers with a fluid stream of information about their students’ performance, which they could interpret, and subsequently modify their instructional behavior as needed. As such, a key mediating variable in this process is instructional differentiation.

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**

The purpose of this paper is to explore the results of an important mediating process in the experiment’s theory of action: instructional differentiation. This paper will present the results of a comparison between treatment and control school teachers on a measure of instructional differentiation.

**Setting:**

This work is part of the Indiana Diagnostic Assessment Tools RCT, years 1 (RCT1) and 2 (RCT2). The experiments occurred in Indiana in 2009-10 and 2010-11.

**Population / Participants / Subjects:**

This study used data obtained from treatment and control school teachers of English language arts (ELA) and mathematics in grades 2 and 5 from RCT1 and RCT2.

**RCT1.** 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 included 45 control and 54 treatment teachers; in grade 5 mathematics, 41 control and 47 treatment teachers.

**RCT2.** In grade 2 ELA, the sample included 39 control and 44 treatment teachers; in mathematics, 31 control and 47 treatment teachers. In grade 5 ELA, the sample was 28 control and 11 treatment teachers; in grade 5 mathematics, 30 control and 11 treatment teachers.

**Intervention / Program / Practice**

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:**

Teachers in control and treatment schools in grades 2 and 5 were asked to complete 16 instructional checklists throughout each academic year in each RCT, 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**

Because of the complexities of the checklist data, our measure of differentiated instruction was constructed using the Rasch model. Each dichotomous item was coded as 1 if at least one student was taught using practices different from other students, 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.

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). Items with a mean square greater than 2.0 were iteratively removed from the measurement system. Reliability estimates in this framework indicate how well the measure separates units (i.e. teachers in time, teachers, and schools) along the latent dimension of differentiation.

A hierarchical linear model, with teachers’ scaled differentiation scores nested in schools in time, was then used to compare treatment and control school levels of instructional differentiation. Standardized mean differences were computed for each subject and grade level in each RCT.

**Findings / Results**

*Reliability*. Our 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 in RCT1 where the estimated log-level reliability estimate was .57 and the estimated school-level reliability estimate was approximately 63.

*Treatment-Control Differences*. When we applied this method to data from the RCT, we found notable differences between the treatment and control group. Table 1 presents the results of these analyses. Figure 1 presents the mean differentiation scores of treatment and control school teachers across the span of the academic year.
In RCT1, grade 5 math and grade 5 ELA teachers in control schools differentiated substantially and statistically significantly more than treatment school teachers in these subjects and grades. Grade 2 ELA and grade 2 math teachers in RCT1 each had substantively large effect size estimates (greater than .10). In RCT2, the largest meaningful difference was observed in grade 5 mathematics. What is interesting about these results is that where meaningful differences in instructional differentiation emerge, they are in the opposite direction of what our theory of action suggested. The magnitude of the differences between treatment and control school teachers is noticeably different between the two trials.

Conclusions:

While the results of this study indicate that teacher checklists can be used to build reliable measures of differentiated instruction, the unexpected direction of the results of our comparisons of treatment and control teachers leave us searching for theoretical explanations.

One explanation may be that teachers’ differentiation is actually suppressed by the feedback they receive about their students as part of the intervention. That is, teachers may view the feedback they receive for the assessment systems as confirmatory, decreasing their propensity to shift their instructional approaches. Another possible explanation is that teachers use the results to decide what to re-teach the whole class, rather than what to chance for individual students. Finally, teachers may differentiate their instruction in a number of ways not captured by the logs. While this may be true, and likely is on some level, we believe that our operational definition of instructional differentiation was broad, allowing ample opportunity for differentiation to occur.
Appendix A. References


### Appendix B. Tables and Figures

Table 1. Treatment and control school differentiation standardized mean differences.

<table>
<thead>
<tr>
<th></th>
<th>Slope Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RCT1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 5 Math (Treatment)</td>
<td>-0.38</td>
<td>0.17</td>
<td>35</td>
<td>-2.16</td>
<td>0.03</td>
</tr>
<tr>
<td>Grade 2 Math (Treatment)</td>
<td>-0.14</td>
<td>0.17</td>
<td>33</td>
<td>-0.83</td>
<td>0.41</td>
</tr>
<tr>
<td>Grade 5 ELA (Treatment)</td>
<td>-0.41</td>
<td>0.16</td>
<td>36</td>
<td>-2.61</td>
<td>0.01</td>
</tr>
<tr>
<td>Grade 2 ELA (Treatment)</td>
<td>-0.12</td>
<td>0.17</td>
<td>34</td>
<td>-0.70</td>
<td>0.49</td>
</tr>
<tr>
<td><strong>RCT2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 5 Math (Treatment)</td>
<td>-0.12</td>
<td>0.28</td>
<td>15</td>
<td>-0.42</td>
<td>0.68</td>
</tr>
<tr>
<td>Grade 2 Math (Treatment)</td>
<td>-0.04</td>
<td>0.23</td>
<td>20</td>
<td>-0.18</td>
<td>0.86</td>
</tr>
<tr>
<td>Grade 5 ELA (Treatment)</td>
<td>0.00</td>
<td>0.30</td>
<td>14</td>
<td>0.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Grade 2 ELA (Treatment)</td>
<td>0.07</td>
<td>0.23</td>
<td>22</td>
<td>0.33</td>
<td>0.74</td>
</tr>
</tbody>
</table>
Lessons Learned from Two Large Scale RCTs: Similarities and Differences

Figure 1. Treatment and control school differentiation means over time.