Journal for the Education of the Gifted
2021, Vol. 44(1) 31–55

DOI: 10.1177/0162353220978304
journals.sagepub.com/home/jeg
SAGE

Validity Evidence for the Revised Classroom Practices Survey: An Instrument to Measure Teachers' Differentiation Practices

Nielsen Pereira[1], Juliana Tay[1], Ophélie Desmet[1], Yukiko Maeda[1], and Marcia Gentry[1]

Abstract
We evaluated the psychometric properties of the Classroom Practices Survey–Revised (CPS-R) when used with students achieving at low, average, and high levels. A total of 739 teachers completed CPS-R for students in their classrooms. Results showed improvement in the reliability of CPS-R across all achievement levels when compared with its previous version. Internal consistency estimates for the four factors were higher for the high-achieving students ($\alpha = .84–.94$) compared with estimates for students who achieve at average ($\alpha = .83–.92$) and low ($\alpha = .81–.90$) levels. Model fit of the data was in the acceptable range across all achievement levels. However, model fit indices for the high-achieving group were slightly better than for the average- and low-achieving groups. Results support the practical value of CPS-R as a tool to assess teachers' use of differentiation strategies.

Keywords
Classroom Practices Survey, differentiated instruction, instrument revision, validity, reliability

Knowledge and practices related to learning and instruction have changed significantly during the past several decades to include increased awareness of the diversity in classrooms (Tomlinson, 2014). Students come from diverse cultural, racial, and/or ethnic backgrounds, and they also differ in development, prior knowledge, ability,

[1]Purdue University, West Lafayette, IN, USA

Corresponding Author:
Nielsen Pereira, Purdue University, 100 N. University St., West Lafayette, IN 47906, USA.
Email: npereira@purdue.edu

learning needs, and interests. In addition, most classroom teachers are challenged with the task of meeting the needs of learners with a wide range of readiness and achievement levels in the same classroom. Failing to take into account differences among students could hinder student learning (Belfi et al., 2012). Therefore, teachers are expected to use differentiated instructional strategies to meet the needs of all students in their classrooms (Belfi et al., 2012; Tomlinson, 2014; VanTassel-Baska & Stambaugh, 2005). Currently, there are few psychometrically sound self-assessment instruments available to assess differentiated classroom instructional strategies as a function of achievement level. The purpose of this study was to examine and evaluate the psychometric properties of the Classroom Practices Survey–Revised (CPS-R) when used with students achieving at low, average, and high levels.

The emphasis on diversity and differentiated instruction is reflected in recent teacher preparation standards and educational policy around the world. In the United States, for example, the Council of Chief State School Officers' (CCSSO, 2013) Interstate Teacher Assessment and Support Consortium (InTASC) Model Core Teaching Standards and Learning Progressions for Teachers include several standards that refer to creating learning environments that recognize individual differences. The National Association for Gifted Children (NAGC) and Council for Exceptional Children (CEC) Teacher Preparation Standards in Gifted and Talented Education (2013) and the NAGC Pre-K-Grade 12 Gifted Programming Standards (NAGC, 2010) also reflect the need for inclusive and differentiated learning environments that recognize individual differences. Moreover, several European education systems are moving toward more inclusion, which inevitably requires more within-classroom differentiation as well (Belfi et al., 2012; Coubergs et al., 2017; Organisation for Economic Co-operation and Development [OECD], 2016; Roiha, 2014).

Nevertheless, teachers often struggle with the implementation of differentiation and do not apply it frequently (Al-Lawati & Hunsaker, 2007; Belfi et al., 2012; Latz et al., 2008; Reis & Renzulli, 2010). This may be because some teachers have a narrow view of differentiation and may think of differentiation as modifying only specific aspects of a classroom or for certain groups (Roiha, 2014). Teachers' use of differentiation has also been linked to beliefs in their ability to differentiate instruction (Suprayogi et al., 2017).

Differentiated Instruction: Theory and Practice

The concept of differentiated instruction is based on the idea that students are inherently different. Differentiated instruction can then be viewed as a philosophy of teaching, a set of teaching principles, and a set of classroom practices that take into account student differences (Tomlinson, 2014). Following the ideas of Renzulli (1986), Archambault et al. (1993) defined differentiated instruction as "modifying the curriculum to meet students' differing learning rates . . . interests, and abilities" (p. ix). In the context of gifted education, differentiated instruction can involve enrichment or acceleration, and it includes experiences that focus on thinking skills, abstract concepts, advanced-level content, interdisciplinary studies, and blending content, process, and product (Archambault et al., 1993, p. x). Kaplan (2009) recommended modifying the depth and complexity of gifted education curricula. Tomlinson and Jarvis (2009) defined differentiated instruction as follows:

> an approach to curriculum and instruction that systematically takes student differences into account in designing opportunities for each student to engage with information and ideas and to develop essential skills. Differentiation provides a framework for responding to differences in students' current and developing levels of readiness, their learning profiles, and their interests, to optimize the match between students and learning opportunities. These three dimensions of student difference can be addressed through adjustments to the content, process, products, and environments of student-learning, and each is justified by a research-based rationale. (p. 599)

Based on our review of research on differentiated instruction and publications on differentiation strategies, it is clear that differentiation is a multifaceted concept and that its use varies depending on the context in or student group with which it is applied. Differentiating instruction often involves adapting the levels of challenge or pace of instruction. For example, a teacher could provide less challenging activities or additional supports to students achieving at lower levels. Alternatively, students achieving at higher levels could move through the curriculum at a faster pace (e.g., Curriculum Compacting; Reis et al., 1992). Other differentiation strategies may involve providing students with choice (e.g., Think-Tac-Toe; Roberts & Inman, 2009) or different activities for students with different levels of readiness (e.g., tiered activities; Adams & Pierce, 2006). Most differentiation strategies involve modifying instruction according to students' readiness levels, learning profiles, and interests as well as varying the content, the process students use, the products students create, and different aspects of the learning environments (Tomlinson & Jarvis, 2009).

One of the most widely used tools for planning differentiated instruction based on students' levels of readiness is The Equalizer (Tomlinson, 2014). The Equalizer allows teachers to vary several aspects of a differentiated activity, such as the level of complexity, level of abstractness, open-endedness, or pace. For example, students achieving at lower levels could work at a slower pace on an activity involving less complex issues than high-achieving students. A high-achieving student, on the contrary, could focus on abstract concepts and on ill-defined and open-ended problems. Regardless of the approach to differentiating instruction used, the goal is usually finding ways to modify instruction to allow every student to make continuous progress and learn something new every day (Roberts & Inman, 2009).

Differentiated Instruction: Empirical Support

A body of research supporting the benefits of differentiated instruction exists. Studies on differentiation strategies have ranged from small studies conducted by one teacher in their own classroom using an action research approach (e.g., Grimes & Stevens, 2009) to cluster-randomized trials involving multiple schools (e.g., Little et al., 2014; McCoach et al., 2014). Deunk et al. (2018) reviewed 21 studies on the cognitive effects of differentiated practices in primary education and found the effects vary depending on the type of practice and the student ability group being studied. For example, homogeneous ability grouping had no effect on students with average and high ability, but small negative effects on students with low ability. Reis et al. (2011) concluded that differentiated instruction in combination with enrichment improved reading achievement. Tieso (2005) examined the effects of differentiated instruction and grouping and found that flexible ability grouping in combination with curriculum differentiation resulted in higher mathematics achievement for average- and high-ability students. Gentry (2013) stated that achievement grouping was beneficial for students of all ability levels because it allows teachers to better address individual students' needs. Research also supports the benefits of grouping gifted students in particular (Brulles et al., 2010; Pierce et al., 2011). Grimes and Stevens (2009) found that clustering students by readiness levels led to improvements in academic performance for low-achieving students and in understanding of mathematics for high-achieving students.

Researchers have also focused on the effects of different types of curricular modifications (e.g., tiered activities or individual choice) on student achievement and understanding. McCoach et al. (2014) found that using tiered mathematics activities had the greatest benefit for high-achieving students in lower achieving schools, but other studies have shown that tiered activities benefit students at various achievement levels (Konstantinou-Katzi et al., 2013; Simpkins et al., 2009).

Teachers' use of differentiated instruction has also been addressed in the literature with studies focusing on how often teachers use differentiated instruction (e.g., Archambault et al., 1993; Moon et al., 1995; Pozas et al., 2020; Suprayogi et al., 2017), factors affecting such use (e.g., King, 2010; Suprayogi et al., 2017), and their perceptions of differentiated instruction (King, 2010; Santangelo & Tomlinson, 2012). Archambault et al. (1993) and Moon et al. (1995) are the most recent publications reporting on how often teachers in the United States differentiate instruction with nationally representative samples. Pozas et al. (2020) conducted a national survey of teachers in Germany and found that tiered assignments and ability grouping were among differentiation strategies most frequently used. Suprayogi et al. (2017) found that teachers in Indonesia implement differentiation to a great extent ($M = 7.31$ on a 10-point scale) and that teachers' self-efficacy beliefs with regard to differentiation and constructivist teaching beliefs as well as classroom size were associated with teachers' use of differentiation. King (2010) found that teacher's knowledge of differentiation was the best predictor of whether teachers differentiated instruction.

Classroom/Teaching Practices Instruments

With growing attention to diversity and emphasis on differentiated instruction, there is increasing demand to learn what actually happens in classrooms. This information on classroom practices can inform professional development and program evaluation (Hora, 2015). Various assessment tools, such as the UTeach Observation Protocol (UTOP; Walkington et al., 2011) and the Inquiring into Science Instruction Observation Protocol (ISIOP; Minner & DeLisi, 2012), have been developed for the purpose of collecting information on classroom practices. However, reflection on action is an important part of the learning process (Kelchtermans, 2009), so self-assessment instruments can play an important role in improving teaching practices. Self-assessment instruments such as the Approaches to Teaching Inventory (ATI; Trigwell & Prosser, 2004) and the Teaching Practices Inventory (Wieman & Gilbert, 2014) aim at assisting teachers in reflecting on their approach to teaching. Nevertheless, few instruments exist for measuring classroom/teaching practices with well-established evidence for reliability and validity of their data (Walkington et al., 2011).

We searched the literature for existing classroom and teaching practice instruments, using Google Scholar and different combinations of one or more of the following search terms: classroom, teaching, practices, observation, instrument, checklist, and protocol. This search revealed three self-assessment instruments and eight observation protocols. Of these, three observation protocols and only one self-assessment instrument had acceptable reliability and validity evidence (see Table 1). Considering the importance of being able to gather information on classroom practices for evaluation purposes, this lack of psychometrically sound instruments for this purpose is concerning. A need exists for instruments that can yield valid and reliable data to help teachers reflect on their classroom practices relating to differentiation. Although the Classroom Practices Survey (CPS: Archambault et al., 1993) has been used by researchers and practitioners, some items are dated, and our previous research (Pereira et al., 2019) identified the need for improvement of the validity evidence of CPS. Therefore, the aim of the study was to update the CPS (Archambault et al., 1993) to fill this gap in the literature.

CPS: Initial Development and Revisions

CPS was originally developed to assess the extent to which teachers differentiate instruction for gifted and talented students in their regular classrooms (Archambault et al., 1993). The original CPS contained 39 Likert-type items with six options (0 = *never* to 5 = *more than once a day*) to measure the frequency with which teachers used a variety of differentiation strategies (i.e., classroom practices) with average- and high-achieving students. The six subscales in CPS were (a) Questioning and thinking (QT), (b) Providing challenges and choices (CC), (c) Reading and writing assignments (RW), (d) Curriculum modifications (CM), (e) Enrichment centers (EC), and (f) Seatwork (SW).

Our research team examined the psychometric properties of the original CPS as limited current construct validity evidence for CPS exists (Pereira et al., 2019). Thus, as a part of a larger research project, responses to the CPS were collected from 648 teachers from six different states, and the data were used in a validation study (Pereira et al., 2019). Specifically, we also examined whether CPS can be used to examine teachers' differentiated instruction practices with low-achieving students in regular classrooms. Results of confirmatory factor analyses (CFAs) indicated that the model fit for the original, six-factor CPS was best for the high-achieving group. However, for the average- and low-achieving groups, the model fit was not at the desired level (Pereira et al., 2019). As a follow-up analysis, two CPS subscales (i.e., EC and SW)

Table 1. Overview of Teaching/Classroom Practices Instruments.

Instrument (author, year)	Data collection method	Reliability and validity evidence
[a]Approaches to Teaching Inventory (Trigwell & Prosser, 2004)	Self-report	Cronbach's alpha for the Conceptual Change/Student-focused approach scale was .73 and for the Information Transmission/Teacher-focused approach scale was .75. Good evidence of statistical validity of the two-factor model, eigen values were > 1.03.
Classroom Observation Protocol for Undergraduate STEM (Smith et al., 2013)	Observation	The interrater reliability was above .90 for all items.
[a]Classroom Practices Inventory (Hyson et al., 1990)	Observation	Cronbach's alpha for the subscales were .92 for Appropriate Program Items, .93 for Inappropriate Program Items, .96 for Total Program, .88 for Emotional Climate, and .96 for Total Appropriateness. The interrater agreement to the same scale point was 64%; within one scale point, it was 97.7%. A four-factor model was fitted, with all factors loading > .50
Classroom Practices Survey (Archambault et al., 1993)	Self-report	Cronbach's alpha for the six factors ranged from .53 (Seatwork) to .83 (Questioning and Thinking). Only exploratory factor analysis results reported in the original studies.
Inquiring into Science Instruction Observation Protocol (Minner & DeLisi, 2012)	Observation	No information found.
Oregon-Teacher Observation Protocol (Wainwright et al., 2003)	Observation	Interrater reliability varied between 29% and 100%; all but two items have less than 50% exact agreement. Within one point, the range of agreement was 57% to 100%.
Perceptions of Teaching Environment Inventory (Prosser & Trigwell, 1997)	Self-report	CFI = .899 to 1.00; TLI = .696 to 1.00; RMSEA = .000 to .154. SRMR = .005 to .022; and Cronbach's alpha = .40 to .61.
[a]Reformed Teaching Observation Protocol (Piburn & Sawada, 2000)	Observation	The reliability estimates for the subscales were .91 for Lesson Design and Implementation, .67 for Content–Propositional Pedagogic Knowledge, .95 for Content–Procedural Pedagogic Knowledge, .91 for Classroom Culture–Communicative Interactions, and .87 for Classroom Culture–Student/Teacher Relationships. Only EFA was reported; factor loadings ranged from 0.37 to 0.86.
Teaching Dimensions Observation Protocol (Hora & Ferrare, 2014)	Observation	Kappa scores for interrater reliability ranged from .62 to .90. Factor analysis was deemed inappropriate.
Teaching Practices Inventory (Wieman & Gilbert, 2014)	Self-report	Traditional reliability and validity checks were not applicable.
UTeach Observation Protocol (Walkington et al., 2011)	Observation	Not reported.

(continued)

Table 1. (continued)

Instrument (author, year)	Data collection method	Reliability and validity evidence
[a]William and Mary Classroom Observation Scale Revised (VanTassel-Baska et al., 2003)	Observation	Cronbach's alpha for the subscales ranged from .67 to .79 for Curriculum Planning and Delivery, .68 to .73 for Accommodations for Individual Differences, .82 to .94 for Problem Solving, .65 to .78 for Critical Thinking Strategies, .77 to .86 for Creative Thinking Strategies, and .83 for Research Strategies. Intraclass coefficient for the content validity was .98.

Note. STEM = science, technology, engineering, and mathematics; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = Standardized Root Mean Squared Residual; EFA = exploratory factor analysis.
[a]These instruments have adequate or acceptable evidence of validity and/or reliability.

were eliminated because they were not aligned with standards for teacher preparation (CCSSO, 2013; NAGC & CEC, 2013) and the NAGC recommended educational practices for students with gifts and talents (NAGC, 2010). In addition, two items (i.e., 23 and 24 in the original CPS) that did not have satisfactory factor loadings were removed. Model fit estimates with 27 items of the original CPS were in the range for acceptable to good model fit for the high-achieving and the average-achieving group and fit was slightly worse for the low-achieving group (Pereira et al., 2019). Findings suggested there was room for further revisions of the instrument to improve the language and add items reflecting current differentiation practices. Thus, additional revisions were completed to improve its measurement properties.

Purpose of the Study

Scale revision that involves the exclusion of the original items and wording changes in items requires revalidation of the instrument because the modifications might affect the interrelationship of existing items. Therefore, the goal of the current study was to examine and evaluate psychometric properties of the 2017 revision of CPS (hereafter referred to as CPS-R) to assess teaching practices employed by instructors for students at different achievement levels (i.e., low-, average-, and high-achieving students). Specific research questions addressed are the following:

Research Question 1: To what extent does CPS-R yield reliable data that can be used to measure teachers' differentiation practices for students with low, average, and high achievement levels? Do the four CPS-R subscales (i.e., QT, CC, RW, and CM) yield acceptable reliability indices across the three achievement levels?

Research Question 2: To what extent does CPS-R yield valid data that can be used to measure teachers' differentiation practices for students with low, average, or high achievement levels? Do the four CPS-R subscales yield valid data across the three achievement levels?

Method

Data used to examine the psychometric properties of CPS-R were collected as part of a larger research project regarding the cluster grouping model (Gentry et al., 2014). One of the purposes of the larger research project was to examine the implementation of grouping and differentiation strategies used in elementary classrooms by teachers for all students, and CPS-R was used to collect data on teachers' use of such strategies. A total of 67 schools from 20 school districts in seven states participated in the project. In the cluster grouping model implemented in the participating schools, students were placed into classes based on their achievement levels to limit the range of student achievement levels in their classrooms. Teachers in the schools participated in the process of placing students in one of the five achievement categories—low, low average, average, above average, or high achieving—following specific guidelines (Gentry et al., 2014). For example, teachers working with high-achieving students did not have any low-achieving students in their classrooms. Teachers completed CPS-R on all achievement groups in their classrooms, so most teachers completed CPS-R on at least two achievement groups (e.g., average- and high-achieving students). The schools were divided into delayed treatment (n = 40) and treatment (n = 27). A delayed-treatment design was used so treatment schools began implementing cluster grouping in 2015 and the delayed-treatment schools served as control schools for the first 2 years of the study, and then, these schools began implementing cluster grouping in 2017.

Preliminary Study

Two rounds of revisions on the original CPS (Archambault et al., 1993) were completed as a preliminary study. First, based on the findings from a previous study (Pereira et al., 2019), the original CPS was revised by changing the verbiage to be more parsimonious and aligned with the structure of other CPS items. For example, Item 28 from the CC subscale, "Provide a different curricular experience by using a more advanced curriculum unit on a teacher-selected topic" in the revised instrument item became "Use a more advanced curriculum unit," which aligned with the other items. The revised instrument with 27 items was then completed by 1,079 teachers from 19 school districts across nine states in 2015. CFA was conducted on the data to evaluate the structural construct validity of the revised instrument.

The results of the CFA did not meet the criteria for adequate model fit, so a second round of revisions was completed with special attention to increasing reliability of the CM subscale, which had lower internal consistency estimates compared with other constructs measured by CPS. Another goal was to improve construct–item alignment with the four-factor structural model specified by recent research (Pereira et al., 2019) and practice on differentiated instruction. The revisions resulted in additional items for all four subscales, and in particular, for the CM subscale. For these new items to reflect current recommended research-based instructional practices, we reviewed the NAGC programming standards (NAGC, 2010) and research-based practices described in recently published books on differentiation practices (e.g., Gentry et al., 2014; Roberts & Inman, 2009; Tomlinson, 2014; Tomlinson & Jarvis, 2009).

Content validity evidence of new items. After studying the existing literature on instructional practices, we created 26 new items across the four subscales and then conducted a survey to obtain input from experts, including 16 researchers and doctoral students in gifted education on the content validity of the new items. Experts were asked to assign each of the new items to one of the four subscales and to indicate their level of confidence that the item represented that subscale and how relevant the item was to the subscale. After examining the survey responses, 13 items (see Appendix A) were selected for inclusion in the updated instrument (i.e., CPS-R). Items were included in the revised instrument if at least nine of the 16 experts selected the subscale for which an item was written, indicated the item was relevant for that subscale, and stated that they were sure of their choice. The final version of CPS-R (see Online Appendix B) consisted of 40 items and four subscales: (a) QT, (b) CC, (c) RW, and (d) CM.

Current Study

Teachers from participating school districts completed CPS-R as pre- and/or posttests in 2017. We used the CPS-R data obtained from teachers from the treatment schools who joined the study in 2015 (i.e., posttreatment data; after 2 years of implementation of cluster grouping) and CPS-R data from teachers in the control schools who joined the study in 2017 (i.e., posttreatment data; before these schools started implementing cluster grouping). One potential issue that could arise from merging two data sets is that teachers' response patterns may shift systematically due to a treatment effect (Schwartz & Sprangers, 2014), and thus, the structural relationship among constructs for the posttest may have changed from that from the pretest (Howard et al., 1979). Preliminary analyses indicated, however, no major group differences existed with regard to reliability estimates and structural models of underlying factors, and the effect of treatment was minimal for the psychometric evaluation of the scale. Thus, the two data sets were merged.

We collected CPS-R data from a total of 739 teachers, and after using a listwise deletion method of handling missing responses on the scale, the sample sizes for the low-, average-, and high-achieving groups were 707 (405 in treatment schools and 302 in delayed-treatment schools), 658 (375 in treatment schools and 283 in delayed-treatment schools), and 661 (377 in treatment schools and 284 in delayed-treatment schools), respectively. The demographics of the teachers in the sample are summarized in Table 2.

Analysis

Descriptive item analysis. Teachers' responses to CPS-R were analyzed by student achievement level (i.e., low, average, and high achievement) because the individual teachers completed CPS for all three achievement levels, and thus, their responses for

Table 2. Demographic Characteristics of Respondents (N = 739).

Demographic characteristic	*n*	%
Gender		
Male	63	8.53
Female	664	89.90
Ethnicity		
American Indian or Alaskan Native	2	0.27
Asian or Pacific Islander	6	0.81
Black	25	3.38
Latinx	20	2.71
Mixed race	5	0.68
White	654	88.50
Teaching experience (years)		
1–5	119	18.36
6–10	143	22.07
11–20	228	35.19
21+	158	24.38

the three groups can be considered dependent. In addition, according to the cluster grouping model, some teachers do not have all levels of student achievement in their classrooms (Gentry et al., 2014) and provided CPS responses only for the student levels they typically had in their classroom. Descriptive statistics, including means and standard deviations, and frequency distribution were computed for each item for each achievement group.

Reliability analysis. Two types of reliability estimates—Cronbach's alpha for internal consistency (α) and the model-based omega (ω) (McDonald, 1999)—were computed for the four factors. McDonald's omega uses factor loadings and estimated error variances via CFA (Brown, 2015) and does not assume equality of the item-factor correlations among items measuring the same factor so that the reliability estimate tends to be more accurate than Cronbach's alpha (Geldhof et al., 2014). Although some researchers suggest that reliability estimates between .70 and .95 are acceptable (e.g., Nunnally, 1994; Tavakol & Dennick, 2011), the specific cutoff values were not specified for the magnitude of reliability as the reliability estimates differ by data characteristics. Using the acceptable range as a guide, reliability estimates were used to judge overall data consistency across items and to identify the items that affected the consistency among the subscales.

Analysis for construct-related validity evidence. The structural validity of the constructs on CPS-R was evaluated with CFA (Brown, 2015) with the mean- and variance-adjusted weighted least squares (WLSMV) estimator because responses of 6-point Likert-type items produce ordinal scale variables. Teachers' responses for each student ability level were evaluated separately. After preliminary analyses with the model with no cross-loadings, we tested a model in which RW6 (i.e., Item 6 of the RW subscale) cross-loaded on the RW and CC subscales and RW7 and RW8 cross-loaded on the CM subscale. The model with cross-loadings of three items was tested because, after reviewing the wording of the items, it was reasonable to consider these items possibly also measured the CC and CM subscales. To identify the model, the factor loading of the first item in each factor was fixed to 1.0 (Brown, 2015) and all factor correlations and residual variances were freely estimated.

Multiple fit indices were used to determine the adequacy of the tested model. More specifically, we employed the chi-square test of statistical significance and evaluated the Tucker–Lewis index (TLI), and comparative fit index (CFI) with the recommended cutoff values of $>.90$ (Bentler, 1992; Byrne, 2013) indicating acceptable fit and $>.95$ (Byrne, 2013; Hu & Bentler, 1999) indicating good fit. We also used the root mean square error of approximation (RMSEA) with 95% confidence interval of ≤ 0.08 indicating acceptable model fit (Browne & Cudeck, 1992) and the weighted root mean square residual (WRMR) of ≤ 1.0 (Yu, 2002). Note that the cutoff value for WRMR of 1 has been often used and generally accepted in the application of structural equation modeling for ordinal and categorical variables. However, recent research indicates that the adequate cutoff value may vary depending on the model being tested (DiStefano et al., 2018). We used Mplus 7 (Muthén & Muthén, 1998–2012) to conduct these analyses.

Results

Descriptive Statistics and Reliability Estimates of CPS-R

Table 3 contains the item means and standard deviations for all CPS-R items for students at the three achievement levels (i.e., high, average, and low achieving) as well as reliability estimates, alpha and omega, for each factor. Reliability estimates were relatively high, with the alpha coefficients ranging from .81 (RW for low-achieving) to .94 (QT for high-achieving), and omega coefficient ranged from .75 (RW for low-achieving) to .92 (QT for high-achieving). No systematic pattern was observed in average item scores by student achievement group. The average item score was relatively high on QT items and ranged from 4.55 to 5.61 on a 6-point scale, indicating that teachers tended to apply the teaching practices related to questioning and thinking on a daily basis. The average items scores for items in the other subscales ranged from 1.13 to 4.84, and the means of CC7 and RW3 were relatively low for all achievement groups, indicating that teachers used these classroom practices (i.e., *sending students to higher grade levels for specific content instruction* and *assigning book reports*) only once a month or less frequently with students who were achieving at lower levels.

Results for Construct-Related Validity

Table 4 includes factor loadings and standard errors for all CPS items for the three achievement levels. As expected from the increase in reliability estimates, the structural validity of the constructs also improved with the 2017 revisions.

Table 3. Means, Standard Deviations, and Reliability Estimates for All CPS Items.

		High (n = 661)				Average (n = 658)				Low (n = 707)			
Factor	Item	M	SD	α	ω	M	SD	α	ω	M	SD	α	ω
QT	QT1	5.01	1.10			5.09	0.96			5.09	0.95		
	QT2	5.13	1.01			5.08	0.93			5.05	0.92		
	QT3	5.27	0.99			5.27	0.90			5.16	0.94		
	QT4	4.77	1.16			4.65	1.10			4.50	1.13		
	QT5	5.56	0.92			5.61	0.81			5.60	0.71		
	QT6	4.74	1.12			4.67	1.06			4.56	1.07		
	QT7	4.66	1.25			4.62	1.20			4.55	1.18		
	QT8	5.09	1.06			5.10	0.98			5.09	0.92		
	QT9	4.83	1.09			4.79	1.02			4.75	0.99		
	QT10	4.77	1.12	.94	.92	4.73	1.07	.92	.91	4.68	1.05	.90	.87
CC	CC1	3.63	1.65			3.55	1.62			3.64	1.63		
	CC2	2.77	1.71			2.56	1.64			2.56	1.69		
	CC3	3.88	1.51			3.71	1.55			3.65	1.55		
	CC4	4.14	1.59			3.27	1.56			2.75	1.60		
	CC5	3.68	1.67			2.89	1.53			2.42	1.50		
	CC6	4.20	1.78			4.16	1.79			4.24	1.74		
	CC7	1.74	1.55			1.48	1.23			1.47	1.21		
	CC8	4.35	1.43			4.34	1.42			4.35	1.40		
	CC9	3.77	1.66			3.57	1.67			3.41	1.70		
	CC10	3.83	1.39			3.77	1.39			3.77	1.36		
	CC11	3.58	1.38	.86	.80	3.55	1.38	.85	.79	3.50	1.37	.83	.76
RW	RW1	2.39	1.27			2.31	1.21			2.24	1.17		
	RW2	2.62	1.28			2.53	1.23			2.52	1.27		
	RW3	2.01	1.21			1.96	1.16			1.90	1.13		
	RW4	3.12	1.33			3.12	1.31			3.09	1.32		
	RW5	3.23	1.40			3.22	1.39			3.15	1.38		
	RW6	3.19	1.51			3.22	1.50			3.20	1.48		
	RW7	4.22	1.32			4.13	1.32			3.93	1.41		
	RW8	4.12	1.42	.84	.80	4.09	1.40	.83	.78	4.07	1.40	.81	.75
CM	CM1	3.40	1.36			3.41	1.33			3.41	1.32		
	CM2	3.52	1.47			3.39	1.42			3.28	1.46		
	CM3	4.05	1.40			3.76	1.44			3.53	1.52		
	CM4	2.92	1.74			2.83	1.74			2.87	1.71		
	CM5	4.69	1.24			4.74	1.18			4.76	1.11		
	CM6	4.38	1.36			4.42	1.31			4.47	1.26		
	CM7	3.96	1.49			3.99	1.47			4.01	1.43		
	CM8	4.80	1.29			4.84	1.25			4.84	1.21		
	CM9	3.91	1.57			3.94	1.55			3.98	1.52		
	CM10	3.72	1.56			3.51	1.53			3.37	1.59		
	CM11	3.76	1.59	.89	.86	3.82	1.55	.87	.83	3.87	1.54	.85	.79

Note. CPS = Classroom Practices Survey; QT = Questioning and Thinking; CC = Providing Challenges and Choices; RW = Reading and Written Assignments; CM = Curriculum Modifications.

Table 4. Model Fit Estimates for the Three Achievement Groups.

Group	*df*	χ^2	RMSEA	90% CI of RMSEA	WRMR	TLI	CFI
High	731	3,469.23	.075	[.073, .078]	2.14	.89	.90
Average	731	3,650.01	.078	[.075, .080]	2.28	.86	.87
Low	731	3,802.74	.077	[.075, .080]	2.41	.85	.86

Note. RMSEA = root mean square error of approximation; CI = confidence interval; WRMR = weighted root mean square residual; TLI = Tucker–Lewis index; CFI = comparative fit index.

Table 5 includes model fit indices, which were highest for the high-achieving group and reached acceptable levels for RMSEA (.075), TLI (.89), and CFI (.90). However, the WRMR index value of 2.1 deviated largely from Yu's (2002) suggested cutoff value of 1 for a good fitting model. Note that WRMR increases as the sample size increases (Muthén & Muthén, 1998–2012) and as the complexity of the tested model increases. In addition, DiStefano et al. (2018) reported unexpected values for WRMR index in empirical studies. DiStefano et al. (2018) found that fit indices improved when the model misspecification and the low correlations between WRMR with other fit indices (CFI, TLI, RMSEA) they tested under mild model misspecification conditions. Therefore, it may be possible that the large WRMR was due to the sample size and high factor loading observed among indicators. The model fit indices for the average-achieving (RMSEA = .078, TLI = .86, CFI = .87, WRMR = 2.28) and low-achieving (RMSEA = .077, TLI = .85, CFI = .86, WRMR = 2.41) groups were slightly worse than fit indices for the high-achieving group, which is a similar pattern observed in our previous study on CPS (Pereira et al., 2019). These results indicate that perhaps CPS-R may be better suited for use with high-achieving students or gifted education programs.

As depicted in Table 6, the CC factor showed high correlation with the RW and CM factors. For the high-achieving group, the factor correlations between CC and RW and between CC and CM were .72 and .79, respectively. Although this provides evidence that these factors are highly related to each other, results of previous studies on CPS (e.g., Archambault et al., 1993; Pereira et al., 2019) and our follow-up analysis with a higher-order factor CFA did not support the existence of a common higher-order factor among these factors. For example, we tested a model with a higher-order factor connecting the CC and CM subscales, but model fit did not reach acceptable levels across the three achievement levels.

The patterns and degrees of factor loadings were similar across different student achievement groups, except for a few items. The factor loadings for each item tended to be the highest for the high-achieving group and the lowest for the low-achieving group; however, the difference between the factor loadings for these two groups was small (i.e., 0.01–0.05) for most of items. One exception, for example, was CC7, which showed the lowest item mean across all items and had the lowest factor loading (0.47) with the high-achieving group and the highest (0.62) with the low-achieving group.

Table 5. Item Factor Loading and Standard Error Estimates for the Three Achievement Groups.

		High		Mid		Low	
Factor	Item	Factor Loading	SE	Factor Loading	SE	Factor Loading	SE
QT	QT1	0.67	0.02	0.65	0.03	0.63	0.03
	QT2	0.83	0.02	0.79	0.02	0.77	0.02
	QT3	0.80	0.02	0.72	0.02	0.64	0.03
	QT4	0.81	0.02	0.80	0.02	0.76	0.02
	QT5	0.74	0.03	0.67	0.03	0.56	0.04
	QT6	0.80	0.02	0.79	0.02	0.74	0.02
	QT7	0.78	0.02	0.77	0.02	0.74	0.02
	QT8	0.84	0.02	0.83	0.02	0.79	0.02
	QT9	0.90	0.01	0.91	0.01	0.90	0.01
	QT10	0.88	0.01	0.85	0.01	0.84	0.01
CC	CC1	0.51	0.03	0.52	0.03	0.45	0.03
	CC2	0.64	0.02	0.66	0.03	0.65	0.03
	CC3	0.71	0.02	0.71	0.02	0.68	0.02
	CC4	0.73	0.02	0.69	0.02	0.66	0.02
	CC5	0.73	0.02	0.69	0.02	0.69	0.02
	CC6	0.54	0.03	0.48	0.03	0.44	0.03
	CC7	0.47	0.05	0.61	0.05	0.62	0.05
	CC8	0.72	0.02	0.69	0.02	0.66	0.03
	CC9	0.66	0.02	0.63	0.03	0.60	0.03
	CC10	0.78	0.02	0.78	0.02	0.77	0.02
	CC11	0.75	0.02	0.76	0.02	0.72	0.02
	RW6	0.36	0.05	0.30	0.05	0.30	0.04
RW	RW1	0.78	0.02	0.82	0.02	0.81	0.02
	RW2	0.76	0.02	0.79	0.02	0.76	0.02
	RW3	0.70	0.03	0.72	0.03	0.72	0.03
	RW4	0.69	0.02	0.64	0.03	0.65	0.02
	RW5	0.78	0.02	0.73	0.02	0.72	0.02
	RW6	0.39	0.04	0.43	0.04	0.41	0.04
	RW7	0.38	0.04	0.42	0.04	0.40	0.04
	RW8	0.43	0.03	0.45	0.03	0.44	0.03
CM	CM1	0.63	0.02	0.61	0.02	0.53	0.03
	CM2	0.67	0.02	0.66	0.02	0.63	0.02
	CM3	0.81	0.02	0.78	0.02	0.73	0.02
	CM4	0.59	0.03	0.55	0.03	0.52	0.03
	CM5	0.82	0.02	0.80	0.02	0.76	0.02
	CM6	0.76	0.02	0.76	0.02	0.75	0.02
	CM7	0.65	0.02	0.64	0.03	0.62	0.03
	CM8	0.71	0.02	0.67	0.03	0.63	0.03
	CM9	0.74	0.02	0.73	0.02	0.69	0.02
	CM10	0.77	0.02	0.75	0.02	0.72	0.02
	CM11	0.67	0.02	0.66	0.02	0.66	0.02
	RW7	0.35	0.04	0.22	0.04	0.16	0.04
	RW8	0.35	0.04	0.29	0.04	0.25	0.04

Note. QT = Questioning and Thinking, CC = Providing Challenges and Choices, RW = Reading and Written Assignments, CM = Curriculum Modifications.

Table 6. Factor Correlations for the Three Achievement Groups.

Achievement group	Factor	QT	CC	RW
High	CC	.59		
	RW	.37	.72	
	CM	.57	.79	.54
Med	CC	.56		
	RW	.33	.67	
	CM	.53	.71	.49
Low	CC	.48		
	RW	.25	.61	
	CM	.42	.67	.42

Note. QT = Questioning and Thinking; CC = Providing Challenges and Choices; RW = Reading and Written Assignments; CM = Curriculum Modifications.

The three cross-loaded items (i.e., RW6, RW7, and RW8) also had generally lower factor loadings across all three achievement groups if compared with the other items.

Discussion

In this study, we presented the process for revising the CPS (Archambault et al., 1993), which had not undergone additional evaluation or revisions until recently. Psychometric evidence provided by our results supports the conclusion that CPS-R is a good option for collecting information on teachers' use of differentiation strategies as it is more up-to-date and has evidence of internal consistency compared with other similar instruments. Researchers have commonly used observation protocols and self-assessment instruments to collect such information. However, several of the observation protocols created to evaluate classroom/teaching practices do not have any information on internal consistency (e.g., Minner & DeLisi, 2012; Walkington et al., 2011) or only have information on interrater reliability (e.g., Smith et al., 2013; Wainwright et al., 2003). Only one of the observation protocols—the Classroom Practices Inventory (Hyson et al., 1990)—found evidence of adequate internal consistency. However, that inventory was also developed in the 1990s and, to our knowledge, has not undergone additional evaluation or revisions since then. Several self-assessment instruments for instructional practices have lower internal consistency estimates than CPS (e.g., Prosser & Trigwell, 1997; Trigwell & Prosser, 2004; VanTassel-Baska et al., 2003) or lack information on reliability (e.g., Wieman & Gilbert, 2014). Compared with the reliability estimates of the original CPS data, which ranged from .53 to .83, CPS-R has improved internal consistency estimates ranging from .81 to .94. This is a step in the right direction for this instrument and an important improvement.

Internal consistency is a prerequisite to evaluate how well measurement models are supported by data; however, it is crucial to evaluate validity evidence, in particular construct-related validity, such as model fit to data, in the process of scale validation (Schmitt, 1996). Results show that model fit for CPS-R is generally in the acceptable range. However, although the revised instrument is an improvement from the original CPS (Archambault et al., 1993) and the most recent version of CPS (Pereira et al., 2019), the construct-related validity evidence for CPS-R did not reach an optimal level, which leaves us several considerations about the instrument. One consideration relates to the gap between how differentiation has been defined in research and practice (including educational standards) and how it is enacted in current classrooms. It is clear that differentiated instruction has become a common practice in most classrooms (Belfi et al., 2012; Coubergs et al., 2017; OECD, 2016; Roiha, 2014). Teachers are frequently expected to modify their curriculum and instruction based on students' readiness levels, learning needs, and interests. However, the conceptual understanding of what differentiation entails is evolving in education (Tomlinson & Jarvis, 2009), and differentiation practices in the classroom are no longer limited to students' readiness levels, learning needs, and interests. As classrooms continue to become more diverse, and the focus and purpose of educational programs continue to change to meet individual needs, the strategies teachers use to address the needs of different students will also likely change. These continuing changes in what differentiated instruction entails affect how to measure constructs related to differentiation. The limited literature base on all specific aspects of differentiation might have contributed to the challenge in developing an instrument that possesses strong construct-related validity. As with any new or evolving concept, additional studies are needed to establish a strong literature base with empirical evidence to guide instrument development.

Related to the potential disconnect between research and practice, not all differentiation practices recommended by experts are captured in CPS-R. Differentiation and classrooms have significantly changed since the 1990s when CPS was originally developed. In the revision process, we reviewed the literature on differentiated instruction to reflect the most recent practices and knowledge on differentiation for item development. However, limited research exists on using differentiation as an approach to curriculum and instruction in all classrooms rather than for students who are struggling or for those with gifts and talents. Furthermore, some teachers continue to have narrow views of differentiated instruction (Roiha, 2014), which may hinder their use of differentiation in the classrooms. For example, some teachers may believe differentiating instruction for students who are struggling may be more important than differentiating instruction for gifted students (Roiha, 2014) or some strategies listed in CPS-R are more frequently used with gifted students. Thus, it is possible that the CPS-R subscales and items do not capture the full range of differentiation strategies used in today's classrooms, especially those strategies used with students achieving at or below average levels. Although we set out to update CPS and made an effort to include items representing differentiation strategies used with students at all achievement levels, it is possible, because the instrument was originally created by gifted education researchers to investigate how often teachers used differentiation with gifted students (Archambault et al., 1993), the instrument would have required even more substantial changes to adequately capture teachers' differentiated practices with all students.

Researchers seeking to use an instrument like CPS-R should be aware of the recent changes in relevant teacher education standards. For instance, the NAGC-CEC Teacher Preparation Standards in Gifted Education (NAGC & CEC, 2013) include expectations for educators working with students with gifts and talents to also take differences in language, culture, and economic status into consideration when designing curricula for those students. The NAGC-CEC standards also mention strategies to develop students' critical and creative thinking. The InTASC Standards, which guide teacher preparation, include additional expectations such as pacing instruction for individual growth rates and making provisions for individual learning needs. As CPS-R is not necessarily fully aligned with these standards, future researchers may incorporate these policy changes in their scale design.

Nevertheless, CPS-R has practical value for teachers to understand and reflect on their own classroom practices. CPS-R can also be used by administrators and researchers interested in understanding how often educators use specific differentiation practices in their classrooms or how teachers' differentiation practices interact with students' performance. For example, an administrator could use CPS-R results to provide optimal professional development opportunities for teachers who implemented certain differentiation strategies less frequently. However, we emphasize that CPS-R should not be used to make high-stakes decisions regarding teachers or their classrooms as that is not the purpose of the instrument. Educators can use CPS-R to reflect on their own differentiation practices so that they can improve their use of the variety of existing differentiation strategies included in the instrument.

Limitations and Future Research

First, the data in this study are self-reported by teachers, which is a limitation. Future research with CPS-R and similar instruments could include direct observation to compare teachers' self-reported differentiation practices with observational data. Second, generalization should be restricted to teachers with similar characteristics and in similar contexts. In this study, most of the teachers were White and female and had more than 6 years of teaching experience, and all teachers were in schools implementing cluster grouping. Future research should include a more diverse group of educators and could focus on other types of gifted education contexts, such as pull-out programs, self-contained gifted classes, and enrichment or acceleration programs. Another limitation is that we tried to keep as many of the original subscales and items as appropriate. However, those subscales and items may not reflect all current differentiation practices as some of the subscales may be too broad (e.g., curriculum modifications) or perhaps emphasize gifted education practices (e.g., providing challenges and choices), which may have caused some of the issues in model fit, especially with the low- and average-achieving groups. Although we made an effort to remove CPS items that did not reflect current classroom practices and strategies and added items using the current knowledge base on differentiation, room for improvement still exists with regard to model fit.

Conclusion

We improved the original CPS and share the updated instrument to measure differentiated instruction in this article. CPS-R is an option for educators collecting data on classroom practices as the validity and reliability evidence for the current version support its use. However, there is still room for improvement, possibly with new additional constructs that reflect current classroom teaching practices. We hope future research on differentiation practices will focus on the theory and practice of differentiated instruction, which could lead to instruments that capture the full range of differentiation practices used in schools. Additional classroom research is needed to enhance understanding of differentiated instruction and its merits. Additional evidence can provide a more solid foundation to create a better instrument. One lesson learned from this project is that it might be harder to revise or fix an existing instrument with outdated items than to create a new one.

Appendix A

Evolution of CPS Items.

Items in first revision (2015 data)	CPS-R items (2017 data)
Questioning and thinking	
1. Teach thinking skills in the regular curriculum	QT1. Teach thinking skills in the regular curriculum
5. Provide questions that encourage reasoning and logical thinking	QT2. Provide questions that encourage reasoning and logical thinking
10. Ask open-ended questions	QT3. Ask open-ended questions
21. Encourage students to ask higher-level questions	QT4. Encourage students to ask higher-level questions
24. Encourage student participation in discussions	QT5. Encourage student participation in discussions
	QT6. Use students' questions to provide depth and complexity
	QT7. Encourage students to think about thinking (i.e., metacognition)
	QT8. Ask follow-up questions to evaluate student thinking
	QT9. Provide opportunities for students to develop critical thinking
	QT10. Provide opportunities for students to develop creative thinking
Providing challenges and choices	
2. Allow students to leave the classroom to work in another location, such as the school library or media center	CC1. Allow students to leave the classroom to work in another location, such as the school library or media center

(continued)

Appendix A (continued)

Providing challenges and choices	
6. Use contracts or management plans for independent study	CC2. Use contracts or management plans for independent study
11. Give time for independent study or small group projects	CC3. Give time for independent study or small group projects
15. Allow students within your classroom to work from a higher grade–level textbook	CC4. Allow students within your classroom to work from a higher grade-level textbook
23. Use a more advanced curriculum unit	CC5. Use a more advanced curriculum unit
16. Group students by ability across classrooms at the same grade level	CC6. Group students by ability across classrooms at the same grade level
22. Send students to higher grade level for specific content area instruction	CC7. Send students to higher grade level for specific content area instruction
25. Consider students' interest in planning instruction	CC8. Consider students' interest in planning instruction
26. Assign programmed or self-instructional materials	CC9. Assign programmed or self-instructional materials
27. Encourage students to engage in long-range projects	Item deleted
17. Establish interest groups which enable students to pursue individual or small group interests	CC10. Provide opportunities for students to develop and pursue their interests
	CC11. Offer students opportunities to select ways of presenting what they learned
Reading and written assignments	
7. Assign reports	RW1. Assign reports
8. Assign projects or other work requiring extended time for students to complete	RW2. Assign projects or other work requiring extended time for students to complete
12. Assign book reports	RW3. Assign book reports
18. Give creative or expository writing assignments with topics selected by teacher	RW4. Give creative or expository writing assignments with topics selected by teacher
19. Give creative or expository writing assignments with topics selected by student	RW5. Give creative or expository writing assignments with topics selected by student
	RW6. Differentiate writing prompts (by interest, achievement, readiness)
3. Assign advanced-level reading	RW7. Assign reading that is slightly to moderately above students' current reading levels
	RW8. Incorporate student choice in the selection of reading and/or writing assignments

(continued)

Appendix A (continued)

Curriculum modifications	
4. Use pretests to determine mastery	CM1. Use pretests to determine mastery
9. Eliminate curricular material that students have mastered	CM2. Eliminate curricular material that students have mastered
20. Help students understand difficult concepts by re-teaching	Item deleted
13. Give different assignments for students who have mastered regular material	CM3. Give different assignments for students who have mastered regular material
14. Assign different homework based on achievement levels	CM4. Assign different homework based on achievement levels
	CM5. Use ongoing assessment strategies
	CM6. Use a variety of assessment formats
	CM7. Use culturally responsive curricula to engage all students
	CM8. Use a balanced assessment system
	CM9. Use technology to differentiate instruction
	CM10. Allow students to bypass content that they already mastered
	CM11. Use tiered lesson plans

Note. CPS-R = Classroom Practices Survey–Revised; QT = questioning and thinking; CC = challenges and choices; RW = reading and writing; CM = curriculum modifications.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: This research was funded by the United States Department of Education, The Javits Gifted and Talented Students Education Program (Award No. S206A140011).

ORCID iD

Nielsen Pereira https://orcid.org/0000-0002-5399-4622

Supplemental Material

Supplemental material for this article is available online.

References

Adams, C., & Pierce, R. (2006). *Differentiating instruction: A practical guide to tiered lessons in the elementary grades*. Prufrock Press.

Al-Lawati, F. A. K., & Hunsaker, S. L. (2007). Differentiation for the gifted in American Islamic schools. *Journal for the Education of the Gifted*, *30*(4), 500–518. https://doi.org/10.4219/jeg-2007-501

Archambault, F. X., Westberg, K. L., Brown, S., Hallmark, B. W., Emmons, C., & Zhang, W. (1993). *Regular classroom practices with gifted students: Results of a national survey of classroom teachers*. The National Research Center on the Gifted and Talented.

Belfi, B., Goos, M., De Fraine, B., & Van Damme, J. (2012). The effect of class composition by gender and ability on secondary school students' school well-being and academic self-concept: A literature review. *Educational Research Review*, *7*(1), 62–74. https://doi.org/10.1016/j.edurev.2011.09.002

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin*, *107*(2), 238-246. https://doi.org/10.1037/0033-2909.112.3.400

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230-258.

Brulles, D., Saunders, R., & Cohn, S. (2010). Improving performance for gifted students in a cluster grouping model. *Journal for the Education of the Gifted*, 34(2), 200–216. http://files.eric.ed.gov/fulltext/EJ910197.pdf

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

Coubergs, C., Struyven, K., Vanthournout, G., & Engels, N. (2017). Measuring teachers' perceptions about differentiated instruction: The DI-Quest instrument and model. *Studies in Educational Evaluation*, *53*, 41–54. https://doi.org/10.1016/j.stueduc.2017.02.004

Council of Chief State School Officers. (2013, April) *Interstate teacher assessment and support consortium InTASC model core teaching standards and learning progressions for teachers 1.0: A resource for ongoing teacher development*. https://ccsso.org/resource-library/intasc-model-core-teaching-standards-and-learning-progressions-teachers-10

Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, *24*(2018), 31–54. https://doi.org/10.1016/j.edurev.2018.02.002

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 453–466. https://doi.org/10.1080/10705511.2017.1390394

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72–91.

Gentry, M. (2013). Cluster grouping. In C. M. Callahan & J. Plucker (Eds.), *Critical issues and practices in gifted education* (2nd ed., pp. 107–115). Prufrock Press.

Gentry, M., Paul, K. A., McIntosh, J., Fugate, C. M., & Jen, E. (2014). *Total school cluster grouping and differentiation: A comprehensive research-based plan for raising student achievement and improving teacher practices*. Prufrock Press.

Grimes, K. J., & Stevens, D. D. (2009). Glass, bug, mud. *Phi Delta Kappan*, *90*, 677–680. https://doi.org/10.1177/003172170909000914

Hora, M. T. (2015). Toward a descriptive science of teaching: How the TDOP illustrates the multidimensional nature of active learning in postsecondary classrooms. *Science Education*, *99*(5), 783–818. https://doi.org/10.1002/sce.21175

Hora, M. T., & Ferrare, J. (2014). *The teaching dimensions observation protocol (TDOP) 2.0*. Wisconsin Center for Education Research, University of Wisconsin-Madison. http://tdop.wceruw.org/

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychology Measurement*, *3*(1), 1–23. https://doi.org/10.1177/014662167900300101

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55.

Hyson, M. C., Hisch-Pasek, K., & Rescorla, L. (1990). The classroom practices inventory: An observation instrument based on NAEYC's guidelines for developmentally appropriate practices for 4- and 5-year-old children. *Early Childhood Research Quarterly*, *5*(4), 475–494. https://doi.org/10.1016/0885-2006(90)90015-S

Kaplan, S. (2009). Layering differentiated curricula for the gifted and talented. In F. A. Karnes & S. M. Bean (Eds.), *Methods and materials for teaching the gifted* (3rd ed., pp. 107–136). Prufrock Press.

Kelchtermans, G. (2009). Who I am in how I teach is the message: Self-understanding, vulnerability, and reflection. *Teachers and Teaching*, *15*(2), 257–272. https://doi.org/10.1080/13540600902875332

King, S. (2010). *Factors associated with inclusive classroom teachers' implementation of differentiated instruction for diverse learners* (Publication No. 3433402) [Doctoral dissertation]. ProQuest Dissertations and Theses Global, Tennessee State University.

Konstantinou-Katzi, P., Tsolaki, E., Meletiou-Mavrotheris, M., & Koutselini, M. (2013). Differentiation of teaching and learning mathematics: An action research study in tertiary education. *International Journal of Mathematical Education in Science and Technology*, *44*(3), 332–349. https://doi.org/10.1080/0020739X.2012.714491

Latz, A. O., Speirs Neumeister, K. L., Adams, C. M., & Pierce, R. L. (2008). Peer coaching to improve classroom differentiation: Perspectives from project CLUE. *Roeper Review*, *31*(1), 27–39. https://doi.org/10.1080/02783190802527356

Little, C. A., McCoach, D. B., & Reis, S. M. (2014). Effects of differentiated reading instruction on student achievement in middle school. *Journal of Advanced Academics*, *25*(4), 384–402. https://doi.org/10.1177/1932202X14549250

McCoach, D. B., Gubbins, E. J., Foreman, J., Rubenstein, L. D., & Rambo-Hernandez, K. E. (2014). Evaluating the efficacy of using predifferentiated and enriched mathematics curricula for Grade 3 students: A multisite cluster-randomized trial. *Gifted Child Quarterly*, *58*(4), 272–286. https://doi.org/10.1177/0016986214547631

McDonald, R. P. (1999). *Test theory: A unified approach*. Lawrence Erlbaum Associates.

Minner, D., & DeLisi, J. (2012). *Inquiring into science instruction observation protocol (ISIOP): Data collection instrument*. Education Development Center.

Moon, T., Tomlinson, C. A., & Callahan, C. (1995). *Academic diversity in the middle school: Results of a national survey of middle school administrators and teachers* (Research Monograph 95124). National Research Center on the Gifted and Talented. https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm95124.pdf

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.).

National Association for Gifted Children. (2010). *2010 pre-K-Grade 12 gifted programming standards*. http://www.nagc.org/sites/default/files/standards/K-12%20programming%20standards.pdf

National Association for Gifted Children, & Council for Exceptional Children. (2013). *NAGC—CEC teacher preparation standards in gifted and talented education*. http://www.nagc.org/sites/default/files/standards/NAGC-%20CEC%20CAEP%20standards%20%282013%20final%29.pdf

Nunnally, J. C. (1994). *Psychometric theory* (3rd ed.). Tata McGraw-Hill Education.

Organisation for Economic Co-operation and Development. (2016). *PISA 2015 results: Policies and practices for successful schools, PISA* (Vol. 2). OECD Publishing. http://dx.doi.org/10.1787/9789264267510-en

Pereira, N., Tay, J., Maeda, Y., & Gentry, M. (2019). Differentiation as measured by the Classroom Practices Survey: A validity study updating the original instrument. *Learning Environments Research*, *22*, 443–460. https://doi.org/10.1007/s10984-019-09284-z

Piburn, M., & Sawada, D. (2000). *Reformed teaching observation protocol (RTOP): Reference manual* [ACEPT technical report]. https://eric.ed.gov/?id=ED447205

Pierce, R., Cassady, J., Adams, C., Neumeister, K., Dixon, F., & Cross, T. (2011). The effects of clustering and curriculum on the development of gifted learners' math achievement. *Journal for the Education of the Gifted*, *34*(4), 569–596. https://doi.org/10.1177/016235321103400403

Pozas, M., Letzel, V., & Schneider, C. (2020). Teachers and differentiated instruction: Exploring differentiation practices to address student diversity. *Journal of Research in Special Educational Needs*, *20*(3), 217–230. https://doi.org/10.1111/1471-3802.12481

Prosser, M., & Trigwell, K. (1997). Relations between the perceptions of teaching and approaches to teaching. *Educational Psychology*, *67*(1), 25–35. https://doi.org/10.1111/j.2044-8279.1997.tb01224.x

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, *48*(2), 462–501. https://doi.org/10.3102/0002831210382891

Reis, S. M., & Renzulli, J. (2010). Is there still a need for gifted education? An examination of current research. *Learning and Individual Differences*, *20*(4), 308–317. https://doi.org/10.1016/j.lindif.2009.10.012

Reis, S. M., Burns, D. E., & Renzulli, J. S. (1992). *Curriculum compacting: The complete guide to modifying the regular curriculum for high ability students*. Prufrock Press.

Renzulli, J. [Ed.]. (1986). *Systems and models for developing programs for the gifted and talented*. Creative Learning Press.

Roberts, J. L., & Inman, T. (2009). *Strategies for differentiating instruction: Best practices for the classroom*. Prufrock Press.

Roiha, A. S. (2014). Teachers' views on differentiation in content and language integrated learning (CLIL): Perceptions, practices and challenges. *Language and Education*, *28*(1), 1–18. https://doi.org/10.1080/09500782.2012.748061

Santangelo, T., & Tomlinson, C. A. (2012). Teacher educators' perceptions and use of differentiated instruction practices: An exploratory investigation. *Action in Teacher Education*, *34*(4), 309–327. https://doi.org/10.1080/01626620.2012.717032

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schwartz, C. E., & Sprangers, M. A. G. (2014). Response shift. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 5542–5547). Springer.

Simpkins, P. M., Mastropieri, M. A., & Scruggs, T. E. (2009). Differentiated curriculum enhancements in inclusive fifth-grade science classes. *Remedial and Special Education*, *30*(5), 300–308. https://doi.org/10.1177/0741932508321011

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *Life Science Education*, *12*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154

Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in the classroom. *Teaching and Teacher Education*, *67*, 291–301. https://doi.org/10.1016/j.tate.2017.06.020

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tieso, C. (2005). The effects of curricular differentiation with between- and within-class grouping. *Journal for the Education for the Gifted*, *29*(1), 60–89. https://doi.org/10.1177/016235320502900104

Tomlinson, C. A. (2014). *Differentiated classroom: Responding to the needs of all learners*. ASCD.

Tomlinson, C. A., & Jarvis, J. (2009). Differentiation: Making curriculum work for all students through responsive planning and instruction. In J. S. Renzulli, E. J. Gubbins, K. S. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems and models for developing programs for the gifted and talented* (pp. 599–628). Creative Learning Press.

Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, *16*(4), 409–424. https://doi.org/10.1007/s10648-004-0007-9

VanTassel-Baska, J., Avery, L., Struck, J., Feng, A., Bracken, B., Drummond, D., & Stambaugh, T. (2003). *The William and Mary Classroom Observation Scales–Revised*. Center for Gifted Education, The College of William and Mary-School of Education. https://education.wm.edu/centers/cfge/resources/instruments/index.php

VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory into Practice*, *44*(3), 211–217. https://doi.org/10.1207/s15430421tip4403_5

Wainwright, C. L., Flick, L. B., & Morrell, P. D. (2003). Development of instruments for assessment of instructional practices in standards-based teaching. *Journal of Mathematics and Science: Collaborative Explorations*, *6*(1), 21–46. https://doi.org/10.25891/NTKY-AX16

Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L., & Marder, M. (2011). *Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program* (UTeach Technical Report 2011-01). UTeach Natural Sciences, University of Texas at Austin. https://utop.uteach.utexas.edu/

Wieman, C., & Gilbert, S. (2014). The Teaching Practices Inventory: A new tool for characterizing college and university teaching in mathematics and science. *Life Science Education*, *13*(3), 552–569. https://doi.org/10.1187/cbe.14-02-0023

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Publication No. 3066425). [Doctoral dissertation, University of California, Los Angeles]. ProQuest Dissertations & Theses Global.

Author Biographies

Nielsen Pereira, PhD, is assistant professor of gifted, creative, and talented studies at Purdue University. His research interests include the design and assessment of learning in varied gifted and talented education contexts, understanding gifted and talented student experiences in talent development programs, and conceptual, contextual, and measurement issues in the identification of gifted and talented populations.

Juliana Tay, PhD, is part of the research team for the Total School Cluster Grouping project, focusing on the evaluation and implementation of the intervention. Her research interests include gifted visual arts learners, issues in identification for visual arts giftedness, and evaluation of gifted programs.

Ophélie Desmet, PhD, is a post-doctoral research fellow at Purdue University. Her research focuses on underachievement and talent development, underrepresentation, and effective classroom practices in gifted education.

Yukiko Maeda, PhD, is an associate professor of educational psychology and research methodology in Department of Educational Studies at Purdue University. She has a strong interest in applied research methodology. In particular, she has expertise in meta-analysis, multilevel modeling, and data-driven decision-making with institutional assessment data.

Marcia Gentry, PhD, professor of educational studies, directs the Gifted Education Research and Resource Institute at Purdue University where she enjoys working with doctoral students and engaging in research. She contributes to the field through service to NAGC and AERA and by writing, reviewing, and presenting research aimed to improve education for children, youth, and teachers.