An Instrument to Measure Teacher Practices to Support Personalized Learning in the Middle Grades

Mark W. Olofson
Center for Education Research and Policy Studies
University of Texas at El Paso
El Paso, TX
molofson@utep.edu

John M. Downes, Carmen Petrick Smith, Life LeGeros, and Penny A. Bishop
Tarrant Institute for Innovative Education
The University of Vermont
Burlington, VT

Abstract

Reforms to support and expand personalized learning increasingly are being introduced in middle schools across the United States. Personalization, as enacted in response to these reforms, encourages teachers to implement many practices that long have been recommended by advocates of middle grades philosophy. To better understand the practices of middle grades teachers working in schools attempting to implement personalized learning, this article presents a survey instrument to measure teacher practices for personalization in the middle grades. The article describes the formulation and initial administrations of the survey to 232 teachers in 2016 and 165 teachers in 2017. Exploratory factor analysis provided evidence for the presence of factors describing practices for personalized assessment, out-of-school learning, whole group learning in a personalized setting, and technology implementation. Confirmatory factor analysis with the follow-up sample provided additional support for this structure. Data from these two separate survey administrations demonstrated high internal consistency and moderate correlation across the groups of practices. Suggestions for future research using the tool are offered. The survey instrument is included as an appendix.

Keywords: personalized learning, middle school, professional development, survey research

Introduction

Proponents and researchers of the middle grades concept (e.g., Association for Middle Level Education, 2012; Jackson, Davis, Abeel, & Bordonaro, 2000; National Middle School
Association, 1982) generally have framed recommended practices for teaching young adolescents within the context of familiar school designs. Learning happens within the school building in a traditional space; classrooms are occupied by students and teachers fulfilling traditionally defined roles; and time is managed in accordance with daily and yearly school schedules. Although many reform initiatives in the middle grades have sought to refine schools and practices to be more responsive to young adolescents (McEwin & Greene, 2011; Smith & McEwin, 2011), these reform efforts have generally operated within the vision of the space, roles, and time that are traditionally found in schools.

More recently, change initiatives in many school districts and state systems in the United States have embraced reforms that move teaching and learning toward more personalized practices (Patrick, Worthen, Frost, & Gentz, 2016). Personalized learning reconsiders the traditional uses of time, space, and roles in the interest of more engaging and successful learning. Although definitions vary, most conceptions of personalized learning involve tailoring curriculum, instruction, and assessment to the interests, abilities, and needs of each individual student (Bray & McClaskey, 2015). Personalized education has shown positive impacts on student achievement in the near term (Pane, Steiner, Baird, & Hamilton, 2015) and proponents argue that empowering students in this way will help them develop into well-rounded adults who can participate in the workforce of tomorrow (Bray & McClaskey, 2015; Patrick et al., 2016). Policymakers are codifying elements of personalized education, increasing the likelihood of the persistence of such reforms (Patrick et al., 2016; Spencer, 2017). The growing personalization movement, consisting of loosely affiliated groups of education professionals and policymakers who push for reforms to shift public education toward the inclusion of personalized approaches to teaching and learning, redefines the context in which recommended middle grades practices may be applied.

Accordingly, the personalization movement calls for a new measure of middle grades teacher practices, in order to understand the evolving methods of those who enact personalized learning. As highlighted in the recent research agenda set by the Middle Level Education Research Special Interest Group (2016), it is important to identify the “common curricular, instructional, and assessment practices of effective middle grades educators” (p. 11). The pedagogical practices of teachers are complex, intersecting with and dependent upon a wide network of influences, including money, time, leadership, structural elements of the school, and personal values (Grubb, 2011). Due to this complexity, studying teaching practices as they occur, rather than the complex system of influences, is likely to provide useful information for educational stakeholders, including reformers, policymakers, parents, school leaders, teachers, and students themselves. Evaluations and measurements of teaching practices need to be multidimensional in order to match the multidimensional nature of the teaching profession (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012).

In response to this need, we developed a survey that measures middle grades teacher practices as they intersect with practices to advance personalized learning. These practices were organized into the categories of whole group learning, customized learning, personalized assessment, out-of-school learning, supportive communities, family engagement, and technology integration to support personalized learning. Many of these categories of practices have been identified as supportive for learners in general (Buck Institute for Education, 2015; Council of Chief State School Officers, 2011; National Youth Leadership Council, 2008), while the categories of supportive communities and family engagement have been recognized as important for meeting the unique needs of young adolescents in particular (National Middle School Association, 1982, 2010). As traditional notions of time, space, and roles are being left behind, the growing movement toward personalization necessitates revisiting these constructs in order to understand current teaching practices in the middle grades.

The purpose of this study was to build and test a survey tool for measuring teaching practices related to personalization in the middle grades. In this paper, we first provide a background sketch of the intersections between middle schools and design elements of personalized learning. We then describe a survey tool for measuring middle grades teacher perceptions of their teaching practices. Next, we report on the structure of the constructs in the survey as determined by the first administration of the instrument, and report the results from a confirmatory analysis using data from the second administration of the survey. Finally, we discuss future uses of the survey and provide the survey instrument. The final version of the survey is presented in Appendix A.

**Perspectives**

**Surveys of middle grades teacher practice**

Surveys allow for individuals to communicate their personal perceptions while generating data that are
interpretable across participants (Dillman, Smyth, & Christian, 2014). Although surveys are but one technique of understanding teacher practices (Darling-Hammond et al., 2012), they are a useful component of multi-method evaluation procedures (Martinez, Schweig, & Goldschmidt, 2016). Surveys of teaching practices can demonstrate acceptable levels of reliability and validity; particularly, at the construct level (Mayer, 1999). For interpretability and coherence, it is important that surveys, along with other teacher evaluative tools, are constructed using guiding frameworks (Darling-Hammond et al., 2012).

Prior surveys of middle grades teachers and leaders represent a line of research on middle schools stretching from the 1960s to the present (Huss & Eastep, 2011). These surveys have explored implementation of the middle grades concept (e.g., Faulkner & Cook, 2006; George, 2007; Huss & Eastep, 2011; McEwin & Greene, 2011), including teaching practices that are characteristic of personalized learning approaches. Results comparing random and highly effective middle schools highlight the importance of teaching and learning strategies that involve young adolescents in their own learning, as highly effective middle schools employ less direct instruction and more cooperative and inquiry-based learning opportunities (McEwin & Greene, 2011). Faulkner and Cook (2006) found that 83–97% of teachers surveyed in Kentucky used teaching strategies tailored to the individual needs of young adolescents; a larger study from multiple states found that 68% of teachers employed these strategies (Huss & Eastep, 2011).

Numerous surveys of teacher practices related to personalization exist. For example, the Constructivist Learning Environment Survey measures four dimensions of teaching and learning that intersect with personalization, including student-centeredness, autonomy, the incorporation of prior knowledge, and negotiated curriculum (Johnson & McClure, 2004; Taylor, Fraser, & Fisher, 1997). However, this instrument and similar instruments were constructed prior to the wide adoption of technology in education that has transformed teaching and learning. The LoTI Digital-Age Survey (Mehta & Hull, 2013; Moersch, 1995) measures technology integration and constructivist teaching practices; however, this survey retains a more teacher-directed vision of teaching and learning than called for by personalization proponents. Finally, surveys that measure teacher beliefs related to personalization and learner-centered classrooms (e.g., Akos, Charles, Orthner, & Cooley, 2011; Woolley, Benjamin, & Woolley, 2004) provide perspective on teacher orientations, but are limited in that they do not measure actual practices.

Given the emergent nature of personalized learning definitions and practices, we have yet to find a survey tool that measures teacher practices situated in this context. The survey in this study extends this line of research on middle schools. It incorporates standards and recommendations for teacher practices from professional organizations advancing personalized learning, and organizes these practices into conceptually distinct categories.

### Design elements of personalized learning

A number of organizations have developed frameworks and recommended teaching practices for personalized learning, which informed the design of this survey. These include frameworks of personalization from the Nellie Mae Education Foundation (2013), the Institute for Personalized Learning (Rickabaugh, 2016), and the Bill & Melinda Gates Foundation (Pane et al., 2015). Additionally, we consulted the teacher competencies for personalized learning identified by Jobs for the Future & the Council of Chief State School Officers (2015) and the InTASC teaching standards (Council of Chief State School Officers, 2011). Finally, effective practices for pedagogies incorporated into personalized learning, such as project-based learning (Buck Institute for Education, 2015), service and community-based learning (National Youth Leadership Council, 2008), and design thinking and learning (Carroll et al., 2010), were included.

Four design elements are common across much of the personalization movement, which are also reflected in recent policy initiatives (e.g., Senate Committee on Education, 2013): 1) flexible pathways that include in-school and out-of-school learning opportunities attuned to the interests and needs of each student; 2) personalized learning plans by which students, their families, and their teachers come to know and plan learning for the needs, interests, and abilities of each child; 3) competency-based graduation requirements and assessment strategies that encompass varied forms of evidence from an array of learning opportunities; and 4) student ownership and agency in their learning. Together, these four elements constitute a system in which students can learn both within and outside the traditional school parameters of roles, space, and time. Students can generate evidence of learning that happens anytime and anywhere, including through blended or online...
learning spaces or in dual enrollment programs. This learning can happen in collaboration with peers or adults within and beyond their school community, and can be guided by goals and timetables for learning that are responsive to their personal needs and interests.

**Teaching practices in personalized learning environments**

In response to the systemic reforms needed for personalized learning, as indicated by the design elements previously, an array of teaching approaches reaches greater significance. The design elements common in personalization, as it is currently emerging, necessitate teachers to develop practices in a number of dimensions. As shown in Figure 1, the relationships between the design elements and the categories of practice are not discrete; rather, these design elements are supported by teacher practices from a common set of dimensions.

Middle grades proponents have consistently promoted many of the same teaching practices advanced by the personalization movement. For instance, negotiating integrative curriculum with a group or class of students has been advanced as a way of generating meaningful, relevant, and student-directed learning (Brodhagan & Gorud, 2012; Springer, 2013), and reflects personalized learning’s emphasis on student ownership and agency. The use of student-led portfolio conferences has promoted personalized assessment, reporting, and family involvement (Thompson & French, 2012), echoing the assessment strategies associated with personalized learning. Service learning and community partnerships have enabled out-of-school, interdisciplinary, and project-based learning (Epstein & Hutchins, 2012; Sanders, 2012; Thompson, 2013), operationalizing the flexible pathways element of personalized learning.

Although these practices illustrate strong alignment between personalization and the middle grades concept, personalized education, which embraces the design principles of flexible pathways, personalized

---

**Figure 1.** Connections among design elements of personalized learning and question categories

© 2018 The Author(s).
planning, competency-based assessment, and increased student ownership require a renewed emphasis on other practices that may conflict with frameworks for middle grades best practices (e.g., Association for Middle Level Education (Ed.), 2012; National Middle School Association, 2003). For example, instruction to support whole group learning is necessarily different in non-personalized and personalized classrooms. While teachers in non-personalized learning classrooms may work to build buy-in and consensus on areas of study, in personalized classrooms such practices may not be central. Additionally, supporting learning that takes place out of the classroom is likely to lead to students being taught by adults who are unacquainted with the unique developmental characteristics of young adolescents. Assessment that provides evidence of competency may not be formative in nature. In a policy setting where personalization is conceptualized with these four design principles, leaving some traditional pedagogies unexamined is insufficient. In the creation of this survey instrument, existing core competencies of teacher work were crossed with the design principles of personalization in order to measure these teacher competencies in this new context.

Although survey instruments to measure middle grades teaching practices and elements of personalized learning exist, there is not, to our knowledge, a survey constructed to measure teaching practices supportive of the current conceptualization of personalized learning in the middle grades. The survey in this study was intended to focus exclusively on those practices deemed critical to personalization, including those that would appear most sensitive to changes in time, space, and roles. The purpose of this study was to develop such a survey based on frameworks for teacher practices in personalized learning environments. Starting from these frameworks, we generated categories and items. We then tested the survey with teachers who work in these personalized environments to determine if the theorized survey structure was represented in their responses. We engaged in a number of iterative processes for development to build a survey of teacher practices that would be useful in a personalized education setting in the middle grades.

**Methods**

**Survey development process**

A multistage, iterative process shaped the structure and content of the survey. The research team selected and collected frameworks from professional organizations related to teacher practices for personalization and middle grades best practices. The team identified similar constructs across these frameworks, and the associated discrete behaviors or practices were pooled. These behaviors served as a foundation for the first round of survey items, which were aligned with the frameworks described previously, yielding 54 items in six categories. These categories included whole group learning, customized learning, personalized assessment, out of school learning, supportive communities, and engaged families. Questions within each category were constructed to measure a range of practices related to the category of teaching practices.

A team of five middle grades professional development providers, who had numerous years working as middle grades teachers and administrators in addition to robust experience leading middle grades professional development (PD), provided feedback on the content and wording of items. While some practices were not evident in the field, other descriptors were determined to be too general, necessitating further elaboration. This feedback process resulted in 59 items in the same six categories.

At this stage, the lead survey researcher conducted cognitive interviews to provide information regarding how survey respondents conceptualized the terms and practices included in the survey (Beatty, 2004). Five teachers who had previously engaged with PD related to middle grades personalization participated in structured interviews where they “thought out loud” while completing the survey (Beatty, 2004; Ericsson & Simon, 1993). In these sessions, the participating teachers voiced their real-time interpretations of the content of the items and their justifications for selecting their answers. The lead researcher, serving as the interviewer, took representative notes and wrote research memos following each session. The survey research team summarized, aggregated, and analyzed the cognitive interview data. This process prompted the deletion of conceptually similar and confusing items and the inclusion of technology as a separate category to increase clarity. This yielded 47 questions in seven categories – the six existing categories and the additional technology for learning category – that made up the experimental version of the survey. Initial iterations of the survey are available on request from the corresponding author.

Following the initial administration of the survey and presentations of the instrument to colleagues regionally and nationally, an additional set of
questions was added to understand teacher practices for cultural responsiveness (see Appendix B). These questions drew heavily from the MLER Research Agenda (Middle Level Education Research Special Interest Group, 2016). These questions were contained only in the second administration of the survey and are included in this report to help facilitate future work with this survey instrument.

**Data collection**

The initial administration of the survey occurred in three settings in 2016. First, the survey was given at 19 schools that serve young adolescents, where all middle grades teachers were invited to participate ($N = 170$). These schools were selected due to their partnership with a regional university-based institute to provide PD focused on technology integration and personalization. A PD provider oversaw the survey administration, which was completed online during a faculty meeting at each school during May-June 2016. The second wave occurred at a weeklong workshop in June 2016 ($N = 46$). Participants were middle grades educators from a mixture of 28 schools that had and had not partnered with the institute. Individuals who had not previously completed the survey were invited to do so. Finally, the survey was administered in August-September 2016 to teachers at three new PD partnership schools ($N = 16$). Again, the administration was overseen by a PD provider and conducted at faculty meetings. Together, these 232 teacher responses made up the “2016 Sample.” Using weighted averages, we can characterize the typical school for teachers in the 2016 sample as a rural middle school housing grades 7–8 with 61 students in each grade level, with nearly 100% of students identifying as White, and 35% of the students classified as qualifying for free or reduced-price lunch.

The second administration of this survey took place in 2017. The survey was again presented to faculty at 15 schools who had partnered with the university-based institute for PD. The survey was administered in May-June 2017, and was again delivered online and overseen by a PD provider. This administration yielded the “2017 Sample” ($N = 165$). The 2017 sample consisted of 89 teachers who were also in the 2016 sample and 76 teachers who were new in 2017. The typical school for teachers in the 2017 sample is very similar to the typical school for the 2016 sample.

**Sample**

Table 1 presents basic demographic information about the 2016 sample ($N = 232$) and the 2017 sample ($N = 165$). The samples are less female (70%, 72%) and have a higher prevalence of master’s degree attainment (69%, 76%) than the national population of teachers (Snyder & Dillow, 2015). A slight majority of teachers in the samples have earned a middle grades-specific teaching endorsement (52%, 52%). Race and ethnicity data were not collected, as the population of teachers in the state is predominantly White and inclusion of such data could cause certain teachers to be identifiable. Teachers in 17 subject areas from 42 schools are represented in the 2016 sample, and 16 subject areas from 15 schools in the 2017 sample. The 2016 sample was used for the structural analyses and the 2017 sample was used to provide confirmatory information about the survey as revised based on the initial analyses. Additionally, the 2017 sample provided initial information about the module of questions added to understand practices for cultural responsiveness.

**Analytical procedures**

The survey was developed based on a set of constructs derived from existing frameworks. Following the collection of the 2016 sample, the data were explored for constructs empirically. As summarized by Henson and Roberts (2006), exploratory factor analysis (EFA) without a predetermined number of factors is appropriate in instrument development. These analyses provide a description of the relationships between actual teacher practices and the proposed constructs, and the relationships among the constructs as they are present in the sample. Although the ratio of participants to items was relatively low (4.9:1), EFA was appropriate due to the level of over-determination of the constructs (Hogarty, 2005). These analyses were conducted using SPSS Version 23 (IBM Corp., 2015). The EFA was conducted using principal components’ extraction with an oblique (direct oblimin) rotation and a delta value of 0. The selection of an oblique rotation was appropriate due to the theoretical moderate to high level of correlation among the constructs. Due to the small sample size and expected correlation between the factors, the empirical Kaiser criterion was utilized to determine the number of factors to extract (Braeken & van Assen, 2016), rather than the practice of retaining all factors with an Eigenvalue greater than 1 or visual analysis of a scree plot. Due to the use of the oblique rotation, the structure matrix from the output was used for interpretive work, as the values in the structure matrix represent the full relationship of the indicator with the factor, including that portion of the variance that is related through a correlated factor (Henson & Roberts, 2006).

The factors and the loadings from the structure matrix were interpreted based on the magnitude of the factor
loadings, the content of the items, and the original constructs. The presence of a salient factor alone was not used as justification to retain the item; similarly, the presence of a factor was not used as justification for the retention of the factor in the final survey (Bandalos & Gerstner, 2016). Items were first sorted by their highest factor loadings. Factors with two or fewer items associated with them were removed from the analysis, along with factors that were unable to be interpreted as aligned with the design elements of personalized learning. Then, items with factor loadings with values $\lambda > 0.4$ on multiple factors that remained in consideration were removed from the set of items, as such salient cross loadings indicate a lack of discriminant interpretation. This process of factor exclusion and item exclusion continued iteratively until a final set of items and factors remained. The remaining factors were interpreted based on their constituent items. Finally, subscale characteristics were computed to provide insight into the nature of these new factors.

A confirmatory factor analysis (CFA) was performed using the data from the 2017 sample. CFA is a type of structural equation modeling that allows for the empirical testing of the relationship between observed data and a theoretical model (Brown, 2015). We tested the ability of the 2017 data to reproduce the relationships implied by the subscale structure that resulted from the previously described EFA. We evaluated the fit of the data to the model using the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis fit index (TLI). This analysis was conducted using MPlus Version 7 (Muthén & Muthén, 1998–2012). Due to the limited number of response categories for each item and the small sample size, we used the unweighted least squares means and variance adjusted estimator in the analysis (Muthén, Muthén, & Asparouhov, 2015).

Missing data were handled using multiple imputation. According to Rubin (1987), multiple imputation yields unbiased parameter estimates and allows for the

### Table 1. Sample Descriptions

|                          | 2016 Sample | %     | 2017 Sample | %     |
|--------------------------|-------------|-------|-------------|-------|
| N (complete)             | 223         | 96%   | 153         | 93%   |
| N (incomplete)           | 9           | 4%    | 12          | 7%    |
| N (total)                | 232         | 100%  | 165         | 100%  |
| Male                     | 65          | 28%   | 40          | 24%   |
| Female                   | 152         | 66%   | 103         | 62%   |
| No Gender Specified      | 15          | 6%    | 22          | 13%   |
| BA highest degree        | 69          | 30%   | 34          | 21%   |
| Master’s Degree or higher| 154         | 66%   | 107         | 65%   |
| No Degree Specified      | 9           | 4%    | 24          | 15%   |
| Middle Level Endorsement | 117         | 50%   | 78          | 47%   |
| No Middle Level Endorsement | 106     | 46%   | 71          | 43%   |
| MLE Not Specified        | 9           | 4%    | 16          | 10%   |
| Number of subject areas represented | 17      |       | 16          |       |
| Number of schools represented | 42      |       | 15          |       |
| Age Range                | 22–65       |       | 21–63       |       |
uncertainty related to parameter estimation to be estimated in a reasonable way. That is, point estimates for parameters are the average of the parameter estimates generated from each imputed data set (Enders, 2010; Rubin, 1987). Associated variability can be determined by combining the within and between data set variability (Enders, 2010). Given relatively recent recommendations regarding multiple imputation that argue for greater numbers of imputations (Graham, Olchowski, & Gilreath, 2007), and the relatively small percentage of missing data in the survey data set, a total of ten data sets were imputed ($m = 10$). The descriptive results, alpha estimates, and correlation coefficients are all based on the pooled estimates. Identical procedures were used for the 2016 and 2017 sample data. For the EFA, one imputed data set was used for reasons of parsimony. As noted by Graham (2009), the use of a single imputed data set for such exploratory analysis is an acceptable procedure. The seventh imputed data set as generated by SPSS was randomly chosen for use in the subsequent EFA. For the CFA, the full information maximum likelihood algorithm native to MPlus was used to estimate parameters based on the data available (Muthén & Muthén, 1998–2012).

Results

This section presents the results from the EFA conducted with data from the 2016 sample, including the dimensionality and deviations from the structure of the survey as originally written. Next, descriptive statistics for the de facto subscales are provided based on this new factor structure. Following this, we present the results from the CFA. Finally, scale characteristics from the 2017 sample data are presented.

Exploratory factor analysis

To evaluate dimensionality, an EFA was conducted using all items. The model used an oblique rotation to allow for the related nature of the constructs under investigation. Analysis of the empirical Kaiser criterion (Bracken & van Assen, 2016) supported the extraction of 10 factors. These factors explained 67.15% of the total variance. Due to the correlation among the factors and the use of the oblique rotation, the structure matrix was used for initial interpretation. The structure matrix presents the total variance shared between the factor and the item, including that which may be due to shared variance between the item and a different factor, and covariance between that second factor and the factor of interest.

The structure matrix is presented in Table 2. Items are sorted by the value of their factor loadings.

As described in the methods section, factors and items were analyzed iteratively, dropping factors with two or fewer unique items and dropping items with salient loadings on more than one retained factor. Initially, factor 4 and factor 9 were dropped due to low item counts. Item PA4 and CL6 were dropped due to cross loadings, leaving 4 items on factor 1, all associated with the personalized assessment construct. For factor 2, items TL3 and EF5 were dropped due to their cross loadings. Although cross loadings for numerous items associated with factor 2 were found on factor 8, simultaneous analysis revealed that factor 8 did not have enough unique items to remain viable. Hence, items from factor 2 that cross loaded onto factor 8 were retained. We identified 8 items for factor 2, all associated with the outside of school learning construct. Factor 3 was identified with items belonging mostly to the whole group learning scale, and items WG7 and WG8 were dropped from the factor due to cross loadings. Factor 5 was dropped due to lack of unique items, and factor 6 was identified with four unique items from the technology for learning scale. The remaining factors were dropped from further analysis due to lack of unique items.

Table 3 presents the results-driven interpretation of the remaining 4 factors, with the items sorted onto the modified constructs. These new constructs are highly similar to the original subscale constructs. The definitions for the personalized assessment construct, the out-of-school learning construct, and the technology for learning construct were left unchanged. The whole group learning construct was further described as related to planning and facilitating whole group learning. The constructs related to supportive communities, customized learning, and engaged families were abandoned, as the presence of unified factors identified by items from this survey was not supported from the data. Table 4 provides short descriptions of the newly identified constructs. These results informed subsequent iterations of the survey while providing insight into the nature of the categories of practices under study. To better inform future research endeavors, the characteristics of these new subscales were investigated. Although this descriptive work employed the same data as the categorization of items into de facto subscales, these results can provide formative information for future research with this survey using the adapted form.

Subscale characteristics

Using the new subscale structures, we calculated descriptive statistics. Table 5 presents the mean score, standard deviation, and alpha value for each
Table 2. Structure Matrix from Exploratory Factor Analysis

| Item  | F1  | F2  | F3  | F4  | F5  | F6  | F7  | F8  | F9  | F10 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| PA2   | .855| −.293| .316| .064| .370| −.326| .269| .364| −.094| .184|
| PA1   | .820| −.302| .157| .029| .361| −.280| .312| .354| −.164| .290|
| PA5   | .807| −.298| .202| .319| .299| −.319| .328| .337| −.316| .239|
| PA3   | .783| −.298| .228| .128| .302| −.346| .276| .432| .070| .141|
| PA4   | .713| −.310| .035| .287| .404| −.299| .409| .322| .011| .094|
| CL6   | .567| −.479| .136| .202| .558| −.097| .381| .536| .020| .016|
| OSL4  | .295| −.805| .326| .346| .213| −.268| .301| .412| −.036| −.010|
| OSL6  | .365| −.803| .182| .277| .233| −.212| .277| .454| −.051| −.046|
| OSL3  | .228| −.798| .237| .257| .347| −.322| .367| .351| .003| −.058|
| OSL5  | .353| −.754| .189| .262| .205| −.289| .358| .448| −.153| −.022|
| OSL1  | .218| −.724| .289| .324| .385| −.194| .364| .342| −.154| −.107|
| SC4   | .279| −.721| .020| .325| .117| −.156| .175| .302| −.096| .216|
| TL3   | .381| −.713| .198| .336| .190| −.456| .369| .401| .024| .102|
| EF5   | .429| −.658| .109| .306| .292| −.113| .151| .625| .111| .166|
| OSL2  | .270| −.636| .177| .070| .085| −.333| .364| .175| −.288| .075|
| SC3   | .383| −.534| .113| .385| .185| −.274| .127| .282| −.486| .386|
| WG3   | .144| −.134| .760| .130| .119| −.345| .292| .163| −.022| .034|
| WG6   | .326| −.286| .742| .340| .171| −.359| .492| .123| −.118| −.149|
| WG4   | .259| −.279| .712| −.031| .063| −.334| .455| .120| −.300| .237|
| WG7   | .409| −.408| .615| .452| .335| −.304| .441| .199| −.014| −.091|
| WG8   | .438| −.451| .555| .115| .231| −.293| .404| .235| .037| .160|
| SC2   | .238| −.042| .510| .355| .154| −.259| .151| .041| −.458| .378|
| SC9   | .229| −.346| .231| .832| .091| −.308| .251| .259| −.212| .214|
| SC8   | .201| −.482| .188| .815| .068| −.240| .220| .326| −.099| .118|
| CL4   | .492| −.257| .222| .097| .744| −.271| .234| .225| −.206| .309|
| CL3   | .372| −.139| .173| −.106| .710| −.271| .287| .186| −.275| .341|
| CL5   | .395| −.277| .071| .232| .707| −.174| .177| .265| −.047| .232|
| CL1   | .492| −.401| .366| .132| .510| −.333| .483| .353| −.288| .017|
| TL1   | .358| −.241| .273| .233| .317| −.770| .348| .371| −.040| .125|

(Continued)
the new scales. Skewness and kurtosis statistics were found to be between $-1$ and $1$ for all frequency distributions, indicating normality. The subscales were evaluated for internal consistency using the imputed data. Alpha values for the subscales were uniformly greater than the cutoff of $\alpha = .7$, indicating an acceptable level of internal consistency (Cronbach & Meehl, 1955; Nunally & Bernstein, 1978). As shown in Table 5, these values ranged from $\alpha = .728$ to $\alpha = .890$. Further analysis of the SPSS output revealed that the deletion of any one item from any of the subscales would not result in an alpha value greater than the alpha value for the full scale.

Mean scores were compared across groups sorted by gender, attainment of a Master’s degree, and attainment of a middle-level endorsement using two-tailed $t$-tests. With the exception of the personalized assessment construct, none of the mean scores for the subscales were significantly different across the groups ($p > .05$). The mean scores for the personalized assessment subscale were significantly different across gendered groups, with individuals who identified as female scoring significantly higher than individuals who identified as male ($t(215) = 2.666, p = .008, d = .273$). However, the population of males in the sample was relatively small, resulting in individual males having greater weight in the results than individual females.
The correlations between the mean scores for the different subscales are presented in Table 6. These values range from $r = .384$ to $r = .514$, indicating a moderate to high level of correlation between the constructs as measured in this survey (Cohen, 1988). All correlations were statistically significant ($p < .05$). This indicates that the teaching practices in these different categories are likely to co-occur in the sample teachers’ classrooms.

**Confirmatory factor analysis**

Using the data from the 2017 sample and the finalized subscale structure from the EFA, we conducted a CFA. We specified four latent variables, one each for whole group learning, personalized assessment, out of school learning, and technology for learning. The associated items followed the structure presented in Table 3. We allowed for residual covariance among the three items from the supportive communities subscale to allow for methodological effects (Brown, 2015). During survey administration, the items appeared on the same screen as each other but are associated with different subscales; specifying residual covariance compensates for the correlation in the error term for these items that would be expected due to this element of the survey design. Finally, following the specifications in the

| Item | Personalized Assessment | Out-of-School Learning | Whole Group Learning | Technology for Learning |
|------|-------------------------|------------------------|----------------------|-------------------------|
| PA2  | x                       |                        |                      |                         |
| PA1  | x                       |                        |                      |                         |
| PA3  | x                       |                        |                      |                         |
| PA5  | x                       |                        |                      |                         |
| OSL4 |                         | x                      |                      |                         |
| OSL6 |                         | x                      |                      |                         |
| OSL3 |                         | x                      |                      |                         |
| OSL5 |                         | x                      |                      |                         |
| OSL1 |                         | x                      |                      |                         |
| SC4  |                         | x                      |                      |                         |
| OSL2 |                         | x                      |                      |                         |
| SC3  |                         | x                      |                      |                         |
| WG3  |                         |                         | x                    |                         |
| WG6  |                         |                         | x                    |                         |
| WG4  |                         |                         | x                    |                         |
| SC2  |                         |                         | x                    |                         |
| TL1  |                         |                         | x                    | x                       |
| TL2  |                         |                         | x                    |                         |
| TL8  |                         |                         | x                    |                         |
| TL5  |                         |                         | x                    |                         |

*Note: Only retained items are shown.*
EFA, we allowed the four latent factors to correlate with each other.

We analyzed the fit statistics to interpret the appropriateness of the model for the experimental data. The RMSEA value was 0.070, below an upper cutoff value of 0.08, demonstrating an adequate fit (Brown, 2015; Browne & Cudeck, 1993). The CFI value was 0.930 and the TLI value was 0.918, both above a lower cutoff value of 0.9 indicating a moderate to good fit (Bentler, 1990; Hu & Bentler, 1999). All items had salient factor loadings (λ > .4). Overall, the results from the CFA indicated that the four-factor theoretical model is a good fit for the empirical data in the 2017 sample, providing additional support for this interpretation.

To better understand the cultural responsiveness subscale, we conducted a CFA following the specifications described earlier with an additional fifth latent variable built from the nine items on that subscale (see Appendix B). All items had salient factor loadings (λ > .4) for the subscale. The fit statistics for this new five-factor model were similar to the four-factor model (RMSEA = 0.063, CFI = 0.922, TLI = 0.913). These results indicate that the five-factor model is a good fit for the empirical data, and provides strong support for the inclusion of the cultural responsiveness items in future administrations of the survey.

**Table 4. Intersections of Survey Constructs and Personalized Teaching**

| Survey Constructs          | Descriptions of Constructs                                                                 |
|----------------------------|-------------------------------------------------------------------------------------------|
| Whole Group Learning       | Practices to plan and facilitate the learning of a class or large group of students at the intersection of personalization and democratic classrooms |
| Personalized Assessment    | Practices to support individual students in self-assessment of learning and performance   |
| Out-of-School Learning     | Practices to identify and support opportunities for learning outside of the traditional school day, beyond the school building, and in collaboration with non-teacher facilitators, that are driven by student interest |
| Technology for Learning    | Practices and norms to support learners in using appropriate technology to enhance all elements of the personalized educative experiences |

**Table 5. Internal Consistency of Revised Subscales**

| Module                        | # of Questions | 2016 |       | 2017 |       |
|-------------------------------|----------------|------|-------|------|-------|
|                               |                | Mean | SD    | α    | Mean  | SD    | α    |
| Whole Group Learning          | 4              | 3.363| .6912 | .728 | 3.687 | .7061 | .728 |
| Personalized Assessment       | 4              | 3.631| .7323 | .866 | 3.700 | .6654 | .826 |
| Out-of-School Learning        | 8              | 2.815| .7894 | .890 | 2.931 | .7132 | .854 |
| Technology for Learning       | 4              | 3.654| .7202 | .760 | 3.766 | .7388 | .776 |
| Cultural Responsiveness*      | 9              | 3.590| .6832 | .891 |

Note: N2016 = 232. N2017 = 165. Statistics computed using non-missing data. * this subscale was only included in the 2017 administration of the survey.
This section provides characteristics related to the provisional subscale for cultural responsiveness, which was only included in the 2017 version of the survey. As shown on Table 5, the values for the means and standard deviations of the subscales stemming from the 2017 sample were similar to the values from the 2016 sample. Skewness and kurtosis values were between −1 and 1 for all subscales. Values for internal consistency ranged from $\alpha = .728$ to $\alpha = .891$, above the commonly used 0.7 cutoff. Using the tools and method developed by Diedenhofen and Musch (2016), we determined that the alpha values were not significantly different ($p < .05$) over the two waves of survey administration.

Similar to the 2016 sample, mean scores were compared across the groups as sorted by gender, achievement of a Master’s degree, and middle-level endorsement status. None of the means for the subscales were significantly different across these groups. Additionally, the correlations between the subscales for the 2017 data were computed. Table 7 presents these correlations. The values range from $r = .293$ to $r = .586$, with all correlations significant ($p < .05$). As with the 2016 sample, these correlations again indicate that the teaching practices that are measured by different subscales are likely to co-occur in the classrooms of the sampled teachers. The $r$ values were compared with the corresponding values from 2016 using Fisher’s $r$-to-$z$ transformation (Howell, 2010; Weaver & Wuensch, 2013). No

Table 6.
2016 Subscale Correlations

|                      | Whole Group Learning | Personalized Assessment | Out-of-School Learning | Technology for Learning |
|----------------------|----------------------|-------------------------|------------------------|------------------------|
| Whole Group Learning | 1                    | .384                    | .387                   | .514                   |
| Personalized Assessment | 1                   | .449                    | .418                   |
| Out-of-School Learning | 1                   | 1                       | .401                   |
| Technology for Learning |                      |                          | 1                      |

Note: All correlations are significant $p < .05$. Analysis conducted with pooled data.

Table 7.
2017 Subscale Correlations

|                      | Whole Group Learning | Personalized Assessment | Out-of-School Learning | Technology for Learning | Cultural Responsiveness |
|----------------------|----------------------|-------------------------|------------------------|------------------------|-------------------------|
| Whole Group Learning | 1                    | .293                    | .498                   | .527                   | .586                    |
| Personalized Assessment | 1                   | .358                    | .466                   | .387                   |
| Out-of-School Learning | 1                   | 1                       | .433                   | .493                   |
| Technology for Learning |                      |                          | 1                      | .509                   |
| Cultural Responsiveness |                      |                          |                        | 1                      |

Note: All correlations are significant $p < .05$. Analysis conducted with pooled data.
significant differences \( p < .05 \) in the values of the correlations were found across the two sample administrations.

**Discussion**

The purpose of this study was to build and test a survey tool for measuring teaching practices related to personalization in the middle grades. The content and structure of the initial survey instrument were informed by numerous frameworks that are used to conceptualize such teacher practices and drive pre-service and in-service teacher development. Using data from the initial survey administration, the dimensionality of the survey was evaluated. Results prompted modifications to the structure of the survey and the conceptualization of the constructs. Based on these changes, a CFA was performed with a follow-up sample, which indicated that the factor structure was a good fit for the data. Finally, descriptive statistics were calculated for the two waves of survey data. The findings provide a strong foundation for the use of this survey instrument to measure the teaching practices of middle grades educators.

In particular, the results from the CFA provide strong evidence as to the presence of these factors in the survey instrument. It is clear that the survey, as modified and presented in Appendix A, contains measures of practices for whole group learning, personalized assessment, out of school learning, and technology integration to support personalization. Although the empirically supported structure eliminated a number of the originally proposed items and subscales, the theoretical grounding for these dimensions and their relationships to the design elements of personalized learning could not be ignored. The dropped subscales included practices to support engaging with families, working with communities, and customizing learning at the individual student level. These categories of practices have been intersected with personalized learning by numerous researchers and advocates (Epstein & Hutchins, 2012; Pane et al., 2015; Sanders, 2012). The fact that this instrument did not draw these categories into focus is not evidence for their lack of relation to personalized learning; rather, it is attributable to the instrument itself. Additions or modifications to this survey in future research could yield items to measure these constructs.

The measures of internal consistency provide evidence for the appropriateness of the measure while also serving as a reference for future research using this survey. The values for Cronbach’s Alpha across all subscales and both waves of the survey were consistently high, and these values were consistent in magnitude across the two waves. Although researcher’s use of alpha has long been criticized for taking the place of dimensional analysis (Cortina, 1993), in this study the value was calculated using subscales that resulted from factor analysis, rather than in the place of such an analysis. This allows alpha to be interpreted as a measure of internal consistency within the subscales. It would be expected that such a measure would be stable across administrations, and, for the two samples in this study, this was the case.

All of these subscales exhibited strong internal consistency, and the mean scores were moderately to highly correlated. That is, the measured practices frequently co-occurred. Although the survey was constructed under the assumption that the practices on the different subscales were related but distinct, the data indicate the potential for considering all of these highly related practices as being part of “good teaching.” As found by previous scholars investigating the implementation of middle grades best practices (Faulkner & Cook, 2006; Huss & Eastep, 2011), there is little discriminant validity when considering these philosophically aligned practices. Some of the participants in the cognitive interviews likewise noted that many of these practices could be considered more generally as “good teaching” (Brodhagan & Gorud, 2012; Springer, 2013). In addition, the practices are highly interdependent. If a teacher conducts practices to support personalization generally, they necessarily incorporate many of these discrete practices in order to meet that larger goal. This indicates the potential limited ability of such a partitioned survey of teacher practices to highlight differences in practices aligned with different subscales.

Although the subscales were moderately correlated, this survey can be useful for a number of purposes. Given the emerging nature of policy for personalization, the relationships among these categories may be different in different policy settings. Comparing the correlations across settings could provide insight into the progress or nature of different approaches to the implementation of personalized learning. Relatedly, interventions such as professional development to support personalized learning may take on different foci given the needs and wants of different schools, districts, or individual teachers. Effective evaluation of such interventions is dependent on alignment between program goals and the instrumentation of assessment (Olofson & Garnett, 2018; Wayne, Yoon, Zhu, Cronen, & Garet, 2008). By
offering a number of subscales, an analysis targeted at narrow goals for growth can be conducted using this survey. Finally, although the aggregate results reveal moderate correlations across all categories, there is variance at the individual level. Personal results may be useful to practitioners as they seek to understand their personal practices.

The results from the EFA, while providing interpretable and empirically supported subscales, also highlighted a limitation common in the field of personalized education research. The lack of an independent factor loading for many items originally in the supportive communities subscale was troublesome, as these items were unique among the larger set in that they engaged with the concept of social and cultural diversity. Presentation of this finding to colleagues prompted the addition of the provisional Cultural Responsiveness items to the 2017 administration, as contained in Appendix B. It is critical for middle school teacher practices to include diverse perspectives, as developing educationally responsive multicultural educational experiences have profound and long-lasting effects on young adolescents’ developing attitudes” (Manning, 1999, p. 87).

The results from the EFA intimated that the teacher practices for personalization might be orthogonal to these practices for including diverse perspectives, which is a finding with profound implications. A decrease in these practices for the sake of promoting personalization could be detrimental to the overall development of students. However, results from the 2017 administration presented an alternative portrait. When culturally responsive practices were measured in a more robust way with nine items, they were correlated to a moderate degree with the other categories of practice. Additionally, the CFA model containing the cultural responsiveness latent factor was a good fit for the data. The relationship between culturally responsive practices and other practices for personalization requires additional research; the survey presented in this paper can aid in such investigations.

The sample and nature of the demographic portion of the survey used in this study resulted in a number of limitations. As noted, the sample was racially homogenous, and reflected a higher level of education than the national population of teachers. Although testing did not indicate differences in subscale scores across educational attainment, use of this survey with a more demographically diverse sample of teachers would necessitate additional analyses of variance across demographic groups. Also, in these waves of administration, the survey did not collect information about the teacher’s years of experience in the classroom. Such a variable could be used to test for variance across novice and veteran teacher groups. Finally, the policy environment of the state in which the sample teachers work is undergoing reform toward personalization. Different states are framing and implementing personalization in different ways (Patrick et al., 2016). The results from this investigation cannot be assumed to hold in settings where the conceptualization of personalized learning is vastly different, or in contexts with more traditional conceptualizations of teaching and learning. An informative follow up study to our initial analysis would be a comparison of CFA results for groups of teachers working in different policy contexts, to observe if the empirically supported structure holds across these settings.

We recommend that future research with this survey should employ the modified version based on the original findings, included in Appendix A. Construct validity can be tested by administering this survey side by side with related instruments, such as the survey used by Faulkner and Cook (2006). Test-retest reliability can be investigated with a sample of teachers receiving no interventions or other support in shifting their practices toward greater personalization. Additionally, this survey can be used to measure the impact or efficacy of interventions with in-service teachers such as PD, further indicating the utility of the tool. It should be noted that a particular challenge we sought to address with the instrument was the highly unpredictable future of personalized learning. We wanted an instrument with a viable life span of ten or more years. This required tracing a fine line between descriptions using familiar and accessible language throughout the instrument while leaving room in items and constructs for the rapid evolution of personalized learning. Although these considerations were made, research using this tool in future policy settings should consider the appropriateness of such language, items, and constructs given the research context.

Recently, the middle-level education research community called on researchers to investigate teaching and learning in middle schools, with particular foci on educator development, curriculum integration, social and emotional learning, and digital technologies (Middle Level Education Research Special Interest Group, 2016). This survey tool could aid in these efforts. As middle schools continue to evolve to incorporate learning that is more
personalized, teacher practices in the middle grades necessarily shift. Drawing from frameworks that consider these shifts in light of the unique nature and needs of young adolescents, this survey serves as a useful descriptor of middle school teacher practices in the categories of planning for and facilitating whole group learning, personalized assessment, out-of-school learning, and technology integration. The authors welcome collaborators in middle grades research to use this tool to work together to better understand teacher practices, in order to help teachers better support and guide young adolescents.

ORCID

Mark W. Olofson @ http://orcid.org/0000-0003-4713-4770

References

Akos, P., Charles, P., Orthner, D., & Cooley, V. (2011). Teacher perspectives on career-relevant curriculum in middle school. RMLE Online, 34(5), 1–9. doi:10.1080/19404476.2011.11462078

Association for Middle Level Education (Ed.). (2012). This we believe in action: Implementing successful middle level schools. Westerville, OH: Association for Middle Level Education.

Bandalos, D. L., & Gerstner, J. J. (2016). Using factor analysis in test construction. In K. Schweizer & C. DiStefano (Eds.), Principles and methods of test construction: Standards and recent advances (pp. 26–51). Boston, MA: Hogrefe Publishing Corporation.

Beatty, P. (2004). The dynamics of cognitive interviewing. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), Methods for testing and evaluating survey questionnaires (pp. 45–66). Hoboken, NJ: John Wiley & Sons, Inc.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107(2), 238–246. doi:10.1037/0033-2909.107.2.238

Braeken, J., & van Assen, M. A. L. M. (2016). An empirical Kaiser criterion. Psychological Methods. doi:10.1037/met0000074

Bray, B., & McClaskey, K. (2015). Make learning personal: The what, who, WOW, where, and why. Thousand Oaks, CA: Corwin, A SAGE Company.

Brodhagen, B., & Gorud, S. (2012). Multiple learning approaches. In Association for Middle Level Education (Ed.), This we believe in action: Implementing successful middle level schools (pp. 63–76). Westerville, OH: Author.

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

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (Vol. 154, pp. 136–162). Newbury Park, CA: Sage.

Buck Institute for Education. (2015). Project based teaching rubric. Novato, CA: Author. Retrieved from http://bie.org/object/document/project_based_teaching_rubric

Carroll, M., Goldman, S., Britos, L., Koh, J., Royalty, A., & Hornstein, M. (2010). Destination, imagination and the fires within: Design thinking in a middle school classroom. International Journal of Art & Design Education, 29(1), 37–53. doi:10.1111/j.1476-8070.2010.01632.x

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78(1), 98–104. doi:10.1037/0021-9010.78.1.98

Council of Chief State School Officers. (2011). Interstate teacher assessment and support consortium (InTASC) model core teaching standards: A resource for state dialogue. Washington, D.C.: Author.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281. doi:10.1037/h0040957

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. The Phi Delta Kappan, 93(6), 8–15. doi:10.1177/003172171209300603

Diedenhofen, B., & Musch, J. (2016). cocor: A web interface and R package for the statistical comparison of Cronbach’s Alpha coefficients. International Journal of Internet Science, 1(1), 51–60.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, phone, mail, and mixed-mode surveys: The tailored design method (4th ed.). Hoboken, NJ: Wiley.

Enders, C. K. (2010). Applied missing data analysis. New York, NY: Guilford Press.

Epstein, J. L., & Hutchins, D. (2012). Family involvement. In Association for Middle Level Education (Ed.), This we believe in action:
Implementing successful middle level schools (pp. 181–198). Westerville, OH: Author.
Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis. Cambridge, MA: MIT press. Retrieved from http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=39233
Faulkner, S., & Cook, C. (2006). Assessing the implementation of the middle school concept: A regional study. Current Issues in Middle Level Education, 1(2), 1–14.
George, P. (2007). Special report: The status of programs in Florida’s middle schools. Clermont, FL: Florida League of Middle Schools.
Graham, J. W. (2009). Missing data analysis: Making it work in the real world. Annual Review of Psychology, 60(1), 549–576. doi:10.1146/annurev.psych.58.110405.085530
Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science, 8(3), 206–213. doi:10.1007/s11121-007-0070-9
Grubb, W. N. (2011). The money myth: School resources, outcomes, and equity. New York, NY: Russell Sage Foundation.
Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. Educational and Psychological Measurement, 66(3), 393–416. doi:10.1177/0013164405282485
Hogarty, K. Y. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. Educational and Psychological Measurement, 65(2), 202–226. doi:10.1177/0013164404267287
Howell, D. C. (2010). Statistical methods for psychology. Belmont, CA: Thomson Wadsworth.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. doi:10.1080/10705519909540118
Huss, J. A., & Eastep, S. (2011). A Tri-State study: Is the middle school movement thriving... or barely surviving?. RMLE Online: Research in Middle Level Education, 34(9), 1–13. doi:10.1080/19404476.2011.11462082
IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.
Jackson, A., Davis, G. A., Abeel, M., & Bordonaro, A. (2000). Turning points 2000: Educating adolescents in the 21st century. New York, NY: Teachers College Press.
Jobs for the Future & the Council of Chief State School Officers. (2015). Educator competencies for personalized, learner-centered teaching. Boston, MA: Jobs for the Future.
Johnson, B., & McClure, R. (2004). Validity and reliability of a shortened, revised version of the Constructivist Learning Environment Survey (CLES). Learning Environments Research, 7(1), 65–80. doi:10.1023/B:LERI.000022279.89075.9F
Manning, M. L. (1999). Developmentally responsive multicultural education for young adolescents. Childhood Education, 76(2), 82–87. doi:10.1080/00094056.2000.10522081
Martinez, J. F., Schweig, J., & Goldschmidt, P. (2016). Approaches for combining multiple measures of teacher performance: Reliability, validity, and implications for evaluation policy. Educational Evaluation and Policy Analysis, 38(4), 738–756. doi:10.3102/0162373716666166
Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data?. Educational Evaluation and Policy Analysis, 21(1), 29–45. doi:10.3102/01623737021001029
McEwin, C. K., & Greene, M. W. (2011). The Status of Programs and Practices in America’s Middle Schools: Results from Two National Studies. Westerville, OH: Association for Middle Level Education. Retrieved from www.amle.org
Mehta, V., & Hull, D. M. (2013). Structural validity of the professional development profile of the LoTi Digital-Age Survey. Journal of Psychoeducational Assessment, 31(1), 61–71. doi:10.1177/0734282912454992
Middle Level Education Research Special Interest Group. (2016). The MLER SIG research agenda. Middle Level Education Research Special Interest Group of the American Educational Research Association. Retrieved from http://mlersig.net/research/mler-sig-research-agenda
Moersch, C. (1995). Levels of technology implementation (LoTi): A framework for measuring classroom technology use. Learning and Leading with Technology, 23, 40.
Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2015). Estimator choices with categorical outcomes. Retrieved from https://www.statmodel.com/download/EstimatorChoices.pdf
Muthén, L. K., & Muthén, B. O. (1998–2012). Mplus user’s guide. (7th ed.). Los Angeles, CA: Muthén & Muthén.
National Middle School Association. (1982). This we believe. Columbus, OH: Author.

National Middle School Association. (2003). This we believe: Successful schools for young adolescents. Westerville, OH: Author.

National Middle School Association (Ed.). (2010). This we believe: Keys to educating young adolescents. Westerville, OH: National Middle School Association.

National Youth Leadership Council. (2008). K-12 service learning standards for quality practice. St. Paul, MN: Author. Retrieved from https://nyc整顿.org/standards/

Nellie Mae Education Foundation. (2013). Putting students at the center: A reference guide. Quincy, MA: Author. Retrieved from http://studentsatthecenterhub.org/wp-content/legacyimg/NMEF_sclreframeweb.pdf

Nunally, J. C., & Bernstein, I. H. (1978). Psychometric theory. New York, NY: McGraw-Hill.

Olofson, M. W., & Garnett, B. R. (2018). Measuring the impact of professional development for student-centred pedagogies: A mixed-methods study. Professional Development in Education, 44(3), 342–355. doi:10.1080/19415257.2017.1347805

Pane, J., Steiner, E., Baird, M., & Hamilton, L. (2015). Continued Progress: Promising Evidence on Personalized Learning. RAND Corporation. Retrieved from http://www.rand.org/pubs/research_reports/RR1365.html

Patrick, S., Worthen, M., Frost, D., & Gentz, S. (2016). Promising state policies for personalized learning. Vienna, VA: International Association for K-12 Online Learning (iNACOL). Retrieved from https://eric.ed.gov/?id=ED567893

Rickabaugh, J. (2016). Tapping the power of personalized learning: A roadmap for school leaders. Alexandria, Virginia: ASCD.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York, NY: Wiley.

Sanders, M. G. (2012). Community and business partners. In Association for Middle Level Education (Ed.), This we believe in action: Implementing successful middle level schools (pp. 199–210). Westerville, OH: Author.

Senate Committee on Education. Act 77. An act relating to encouraging flexible pathways to secondary school completion., Pub. L. No. S130 (2013). Retrieved from http://www.leg.state.vt.us/docs/2014Acts/ACT077.pdf

Smith, T. W., & McEwin, C. K. (2011). The legacy of middle school leaders: In their own words. Charlotte, NC: Information Age Pub.

Snyder, T. D., & Dillow, S. A. (2015). Digest of education statistics 2013 (No. NCES 2015-011). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from ERIC Database. (ED483072)

Spencer, K. (2017, August 11). A new kind of classroom: No grades, no failing, no hurry. The New York Times, pp. 20–21.

Springer, M. (2013). Charting the course of curriculum integration. In P. G. Andrews (Ed.), Research to guide practice in middle grades education (pp. 187–216). Westerville, OH: Association for Middle Level Education.

Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. International Journal of Educational Research, 27(4), 293–302. doi:10.1016/S0883-0355(97)90011-2

Thompson, K. (2013). Service learning: Connecting curriculum to community. In P. G. Andrews (Ed.), Research to guide practice in middle grades education (pp. 247–266). Westerville, OH: Association for Middle Level Education.

Thompson, S. C., & French, D. (2012). Varied assessments. In Association for Middle Level Education (Ed.), This we believe in action: Implementing successful middle level schools (pp. 63–76). Westerville, OH: Author.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. Educational Researcher, 37(8), 469–479. doi:10.3102/0013189X08327154

Weaver, B., & Wuensch, K. L. (2013). SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. Behavior Research Methods, 45(3), 880–895. doi:10.3758/s13428-012-0289-7

Woolley, S. L., Benjamin, W.-J.-J., & Woolley, A. W. (2004). Construct validity of a self-report measure of teacher beliefs related to constructivist and traditional approaches to teaching and learning. Educational and Psychological Measurement, 64(2), 319–331. doi:10.1177/0013164403261189
Appendix A: Teacher Practices for Personalization Survey

This appendix contains the final survey questions. All questions were answered using a 5-point Likert Scale. The descriptors provided are:

1. This is a very poor description of my practice.
2. This is a poor description of my practice.
3. This is neither a good not a poor description of my practice.
4. This is a good description of my practice.
5. This is a very good description of my practice.

Whole Group Learning

*Practices to plan and facilitate the learning of a class or large group of students at the intersection of personalization and democratic classrooms*

| I create opportunities for classes or large groups of students to… |
| --- |
| WG1: engage in projects that are interdisciplinary |
| WG2: engage in projects that extend over several weeks or months |
| WG3: identify personal, local or global issues to investigate |
| I create opportunities for students to… |
| WG4: ensure students develop the skills needed for successful collaborative group work |

Personalized Assessment

*Practices to support individual students in self-assessment of learning and performance*

| I create opportunities for students to… |
| --- |
| PA1: identify their own strengths and needs |
| PA2: set meaningful goals |
| PA3: document evidence of progress toward their goals |
| PA4: self-assess (e.g. reflect, analyze) progress to adjust their plans for learning |

Out-of-School Learning

*The current emphasis on personalization, flexible pathways and the power of technology create new opportunities for learning outside of the traditional school day and building.*

| I help students… |
| --- |
| OSL1: plan out-of-school learning related to their interests |
| OSL2: access instructional materials from outside of the classroom |
| OSL3: find out-of-school collaborators (peer or adult) with similar learning goals or interests |
| OSL4: work with out-of-school facilitators (e.g., mentors, community members, afterschool providers, coaches, personal learning networks) |

(Continued)
OSL5: use evidence from out-of-school learning as evidence of academic progress

OSL6: receive assessment feedback on their learning from out-of-school experts

I...

OSL7: ensure students develop the skills needed to learn successfully in out-of-school settings

OSL8: collaborate with out-of-school facilitators to create a safe, positive learning climate for my students (e.g., setting norms with other educators, learners, and adult collaborators).

Technology for Learning

*Practices and norms to support learners in using appropriate technology to enhance all elements of the personalized educative experiences.*

create opportunities for students to use technology to...

TL1: pursue their personal plan for learning

TL2: extend their learning beyond the classroom

TL3: manage their project work

TL4: generate evidence of proficiency in multiple ways (e.g., videos, podcasts, dynamic presentations)

Appendix B: Provisional Questions

This appendix contains provisional questions related to culturally responsive practices. These questions were included only in the 2017 administration of the survey. All questions were answered using a 5-point Likert Scale. Identical to the main survey, the descriptors provided are:

1. This is a very poor description of my practice.
2. This is a poor description of my practice.
3. This is neither a good not a poor description of my practice.
4. This is a good description of my practice.
5. This is a very good description of my practice.

Cultural Responsiveness

*Educators can initiate and support conversations about ethnicity in their classrooms.*

C1: provide opportunities for students to engage with people from other social, racial, linguistic, and cultural backgrounds.

C2: attend to learners’ personal, family, and community experiences and cultural norms when planning learning opportunities.

C3: honor multiple perspectives in class discussions, including those stemming from learners’ personal, family, and community experiences and cultural norms.

(Continued)
C4: bring multiple perspectives to class discussions, including the experiences and cultural norms of communities and individuals other than those represented in my classroom.

C5: provide students with opportunities to demonstrate their own concerns about how they are treated by others in school and in the community.

C6: initiate conversations about ethnicity, race, and culture in my classroom.

C7: work to deepen my understanding of my own frames of reference (e.g., culture, gender, language, abilities, and ways of knowing) and my own potential biases in these frames.

C8: identify and discuss with my colleagues the cultural biases and potential inequity inherent in systems present within our school and community.

C9: provide opportunities for learners to explore all aspects of their identities, including race, culture, and ethnicity.