The Development of a Predictive Opportunity-to-Learn Model Using Machine Learning

2018-2020 National Association for Music Education & Society for Research in Music Education Research Grant

Brian C. Wesolowski

November 1, 2020

National Association for Music Education

"Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think artificial intelligence will transform in the next several years."

Andrew Ng
Associate Professor, Stanford University
Founder, Google Brain Deep Learning Project

Questions related to this research study may be addressed to:

Brian C. Wesolowski, Ph.D.
University of Georgia
Hugh Hodgson School of Music
250 River Road
Athens, GA 30602

Thank you to members of my Summer 2017 graduate seminar for their hard work on the original conception of the OTL rating scale: Mia Athanas, Chris Carr, Krissi Davis, Jacob Fortner, Kalem Graham, Brian Parido, and Jonathan Thompson.

Cover image: © https://unsplash.com/@sharonmccutcheon

Table of Contents

Tables and Figures

Abstract

Opportunity-to-learn (OTL) is the consideration of all conditions or circumstances within schools and classrooms that promote fair and inclusive learning for all students. Educational research suggests that students' differences in academic achievement are not considered in the context of OTL and most districts bypass OTL variables in the analysis of student achievement data. Furthermore, lack of OTL consideration can hamper teachers' instructional practices while also undermining the quality of teaching and learning occurring in the classroom. Currently, the field of music education does not have a mechanism to identify variability in opportunity-to-learn across music classrooms nor does it have a mechanism to provide meaningful guidance and assistance based upon this variability. From instructional, political, sociological, and professional development perspectives, there is a clear need to better understand the effects of opportunity-to-learn variables on music teaching and learning. The purpose of this study was to develop a predictive opportunity-to-learn model for secondary-level instrumental music programs based upon the 2014 Opportunity-to-Learn Standards for Music Instruction. In order to accomplish this task, three aims were addressed.

Aim 1. The first aim of this study was to examine the quality of opportunity-to-learn in secondary level music performance classrooms through the development and validation of an opportunity-to-learn self-report rating scale based upon the 2014 Opportunity-to-Learn Standards for Music Instruction. The research questions that guided the first aim include:

1. What is the psychometric quality (i.e., validity, reliability, and precision) of the rating scale used to measure opportunity-to-learn?
2. How well do the survey items and domains fit the expectations of the measurement model?
3. How do the survey items and domains vary in their ability for respondents to positively agree with them?

Aim 2. The second aim of this study was to identify typologies of opportunity-to-learn using an unsupervised machine learning approach. The research questions that guided the second aim include:

1. Do meaningful opportunity-to-learn typologies exist based upon systematic differential item functioning (item-by-respondent) bias indices?
2. What are the predominant characteristics of the opportunity-to-learn typologies?

Aim 3. The third aim of this study was to build a Random Forest model in order to predict opportunity-to-learn classifications based upon systematic differential item functioning (respondent-by-item) bias indices. The research questions that guided the third aim include:

1. How accurately can a Random Forest model classify music programs into each of their respective opportunity-to-learn typologies?
2. What adjustments to the hyperparameters of the Random Forest model can be made to improve the model's error rate?

3. What are the most important survey items for predicting the opportunity-to-learn typologies of music programs?

The overall goal was to create a model that can accurately and efficiently map an opportunity-to-learn classification to a newly completed survey response.

Overall, the opportunity-to-learn measure exhibited acceptable psychometric properties, indicating that the model's estimated measures for respondents and survey items can be meaningfully interpreted. Results of the analysis suggest strong predictive validity as indicated by statistically significant results and high separation of reliability for the respondents and strong construct validity as indicated by statistically significant results and high separation of reliability for the survey items. The reliability of separation for respondent measures (interpreted similarly to Cronbach's alpha coefficient) was 0.93, indicating that the survey items clearly discriminated between respondents with varying levels of opportunity-to-learn.

Analysis of the residuals indicated that the model estimates were reasonable summaries of music educator responses to the opportunity-to-learn survey items. Additionally, all domains of the survey reasonably fit the expectations of the measurement model, suggesting that all domain-level inferences can be meaningfully interpreted. The rank ordering from the hardest domain for respondents to positively agree with to the easiest domain for respondents to positively agree with included: (a) staffing, (b) materials and equipment, (c) facilities, (d) curriculum, and (e) scheduling.

Based on the respondent-by-item bias indices, a three-cluster solution was identified as having a meaningful structure toward defining opportunity-to-learn typologies. The characteristics of Cluster 1 (34.53% of respondents) included strong depth and breadth of curriculum, strong considerations for students with disabilities, a strong focus on professional development, strong administrative considerations for scheduling music classes, but insufficient funding for equipment. The characteristics of Cluster 2 (32.43% of respondents) included strong funding for materials and equipment but a lack of staffing, staffing qualifications, and staff development, concerns with scheduling and special education, and teaching load concerns. The characteristics of Cluster 3 (33.03% of respondents) included sufficient funding for music literature, sufficient facilities, but a lack of funding for purchasing and maintaining instruments, concerns toward staffing, and a lack of depth and breadth in curricular offerings.

The initial model accuracy for correctly predicting respondents' opportunity-to-learn typologies based the respondent-by-item bias indices was approximately 87%. After adjusting the hyperparameters of the initial model specifications, the optimized model demonstrated an improved accuracy of approximately 89%. Overall, five items demonstrated a significantly weighted importance toward correctly classifying respondents into one of the three opportunity-to-learn classifications: Item 1.11, Item 3.18, Item 3.06, Item 1.03, and Item 3.11.

Future research and action steps are addressed toward better understanding and improving opportunity-to-learn conditions in secondary-level instrumental music classrooms.

Opportunity-to-Learn and Educational Policy

In 1989, the President's Education Summit with the Nation's Governors (Vinovskis, 1999) authorized legislation to (a) advise on the desirability and feasibility of national standards and tests, and (b) recommend long-term policies, structures, and mechanisms for setting voluntary education standards and planning an appropriate system of tests (Education Council Act, 1991). Acting upon this new legislation, the National Education Commission and the National Education Goals Panel charged the newly assembled National Council on Education Standards and Testing to engage education experts and the public-at-large on the desirability and feasibility of national standards and to provide recommendations toward long-term policies, structures, and mechanisms for setting voluntary national standards.

A result of this year-long work was a 1992 report, *Raising Standards for American Education*, that called for "school delivery standards" as an important means toward improving educational structures (National Council on Education Standards and Testing, 1992). According to the report, "school delivery standards:"

> … should set out criteria to enable local and state educators and policymakers, parents, and the public to assess the quality of a school's capacity and performance in educating their students in the challenging subject matter set out by the content standards. School delivery standards should provide a metric for determining whether a school "delivers" to students the opportunity to learn the material in the content standards. (p. 72)

Two years later, the "school delivery standards" manifested more concretely as Opportunity-to-Learn Standards in the 1994 *Goals 2000: Educate America Act*, where coincidently, arts education was written into federal law for the first time. In particular, the statutes in Goals 2000 that highlighted important developmental considerations for opportunity-to-learn standards included the following:

> (a) the quality and availability to all students of curricula, instructional materials, and technologies . . ; (b) the capability of teachers to provide high-quality instruction to meet diverse learning needs in each content area to all students; (c) the extent to which teachers, principals, and administrators have ready and continuing access to professional development . . ; (d) the extent to which curriculum, instructional practices, and assessment are aligned to voluntary national content standards; [and] (e) the extent to which school facilities provide a safe and secure environment for learning and instruction and have the requisite libraries, laboratories, and other resources necessary to provide an opportunity to learn. (108 U.S. Statutes 144)

In response to *Goals 2000*, the Music Educators National Conference (MENC) published the 1994 Opportunity-to-Learn (OTL) Standards for Music Instruction, which included a "comprehensive set of recommendations concerning the types and levels of support necessary to achieve the

the national standards" (Music Educators National Conference, 1994, pp. vi-vii). The OTL Standards for Music Instruction, aligned to the 1994 National Standards for Arts Education, highlighted four strands (e.g., curriculum and scheduling, staffing, materials and equipment, and facilities) across four age groups: (a) prekindergarten and kindergarten (ages 2-5), (b) elementary (grades 1-5 or 1-6), (c) middle school and junior high school, and (d) high school. The OTL Standards for Music Technology (Music Educators National Conference, 1999) were later published as a follow-up addendum to the 1994 OTL Standards for Music Instruction with specifications for curriculum and schooling, staffing, equipment, materials and software, and facilities across K-12 age groups.

The preface of the 1994 OTL Standards for Music Instruction acknowledges that the standards may not necessarily be met by most music programs across the country due to "varying circumstances, practices, and traditions" (p. vii). There are several reasons for this variability addressed in research literature. Morrison (1999) makes a convincing argument that from a music-making perspective, this variability across music programs is an artifact of chronological, stylistic, cultural, and geographic frameworks within the music classroom. From an assessment perspective, Lehman (2014) argues that this variability is an artifact of a lack of standardized curricula. From a psychometric perspective, studies suggest that performance standards naturally vary from district to district and from state to state (U.S. Department of Education, 2010). From a political perspective, support for the arts and more specifically, music, also varies across state and district boundaries (Shuler, Brophy, Sabol, McGreevy-Nichols & Schuttler, 2016). Because of this innate variability across music classrooms, the meeting of OTL Standards for Music Instruction allows for interpretative, judgmental decisions by the most appropriate stakeholders of the music programs who have the most knowledge of the relevant conditions and circumstances (Music Educators National Conference, 1994, pp. vi-vii).

The Council of Music Program Leaders drafted the National Association for Music Education's most current version of the Opportunity-to-Learn Standards (NAfME, 2014) in order to "identify the resources that need to be in place so that teachers, schools, and school districts can give students a meaningful chance to achieve at the levels spelled out in the Core Arts Music Standards" (n.p.). More specifically, the OTL standards are:

> … considered guidance on the Curriculum and Scheduling, Staffing, Materials and Equipment, and Facilities that must be in place if the promise inherent in the Core Music Standards is to be realized – that all American students must have the opportunity to achieve music literacy. (n.p.)

The 2014 Opportunity-to-Learn Standards for Music Instruction bare similar characteristics to the 1994 Opportunity-to-Learn Standards for Music Instruction with the same strands including curriculum and scheduling, staffing, materials and equipment, and facilities. However, the standards were updated to evaluate students across eight distinct age and content groups in order to align with the eight strands of the 2014 Core Arts Standards (All Grades – All Content Areas; PreK-2 General Music; Grade 3-5 General Music; Grade 6-8 and all Secondary General Music; Elementary and Secondary Grades; Ensembles; Composition/Theory; Elementary and Secondary Grades Guitar/Keyboard/Harmonizing Instruments; and Technology).

Need for the Study

Broadly defined, opportunity-to-learn (OTL) is the consideration of all conditions or circumstances within schools and classrooms that promote fair and inclusive learning for all students (Schmidt, 2009). OTL is most often evaluated from an inclusion perspective, which includes the environmental factors that affect learning, or "the structures and processes that define everyday life in schools" (Valli, Cooper, & Frankes, 1997, pg. 254). Winfield and Woodward (1994) identified a list of the five most frequently cited factors in educational research literature that affect students' overall quality of education within the school environment: (a) curriculum, (b) instructional quality, (c) time, (d) resources, and (e) school conditions.

Although research on the Opportunity-to-Learn Standards for Music Instruction themselves is scarce in the field of music education, research related to Winfield and Woodward's five areas in the context of music teaching and learning is pervasive and frequently cited both from practitioner-based and research-based perspectives. In music education research, curricular considerations can include discussions of multiculturalism (Kang, 2016), variability in ensembles within the curriculum itself (Hasket, 2016), considerations towards teaching specific skills (Menard, 2015) and considerations towards interdisciplinary subject relationships (Rogers, 2016; Sotiropoulou-Zormpala, 2016). Research regarding instructional quality often contains questions regarding teacher expertise (Allsup, 2015), licensure and certification practice (May, Willie, Worthen & Pehrson, 2017), teacher evaluation and effectiveness (Shaw, 2016), pedagogy (Crawford, 2017), professional development (Bautista, Yau, & Wong, 2017), and assessment practice (Russell & Austin, 2010). Time considerations include time allocation (Moore, Brotons, & Jacobi-Kama, 2002) as well as scheduling concerns (Baker, 2009; Latten, 1998). Concerns for resources often include considerations towards performance spaces (Blauert & Raake, 2015), parental involvement (Briscoe, 2016; Zdzinski, 2013), technology use (Wise, Greenwood, & Davis, 2011), and advocacy (West, 2012). Examples of research regarding school conditions include methods for educational reform (May & Brenner, 2016), social justice (Salvador & Kelly-McHale, 2017), high-stakes testing (Baker, 2012), school culture (Morrison, 2001), culture bias (Abril, 2009; Kruse, 2016), and culturally responsive teaching (Schmidt & Smith, 2017).

The independent investigations of each of the considerations described by Winfield and Woodward (1994) in the context of music teaching and learning is clearly beneficial towards understanding their unique impact on music programs. However, there is little understanding of how these opportunity-to-learn variables function together within and across music programs.

The Opportunity-to-Learn Standards for Music Instruction (1994) indicate that:

> Both practice and history support the belief that there is a high correlation between effective student learning in music and the existence of the favorable conditions specified in the opportunity-to-learn standards. The correlation is clear, although a cause-and-effect relationship has yet to be documented through research. The experience of generations of music teachers confirms that students are more likely to learn if the specifications stated in the opportunity-to-learn standards are met. (p. vii)

Educational research suggests that students' differences in academic achievement are not considered in the context of OTL and most districts bypass OTL variables in the analysis of student achievement data (Stevens & Grymes, 1993). Furthermore, lack of OTL consideration can hamper teachers' instructional practices while also undermining the quality of teaching and learning occurring in the classroom (Heafner & Fitchett, 2015). Currently, the field of music education does not have a mechanism to identify variability in opportunity-to-learn across music classrooms nor does it have a mechanism to provide meaningful guidance and assistance based upon this variability. From instructional, political, sociological, and professional development perspectives, there is a clear need to better understand the effects of opportunity-to-learn variables on music teaching and learning.

Purpose, Aims, and Research Questions

The purpose of this study was to develop a predictive opportunity-to-learn model for secondary-level instrumental music programs based upon the 2014 Opportunity-to-Learn Standards for Music Instruction. The overall goal was to create a model that can accurately and efficiently map an opportunity-to-learn classification to a newly completed survey response. In order to accomplish this goal, three aims were addressed.

Aim 1. The first aim of this study was to examine the quality of opportunity-to-learn in secondary level music performance classrooms through the development and validation of an opportunity-to-learn self-report rating scale based upon the 2014 Opportunity-to-Learn Standards for Music Instruction. The research questions that guided the first aim include:

1. What is the psychometric quality (i.e., validity, reliability, and precision) of the rating scale used to measure opportunity-to-learn?
2. How well do the survey items and domains fit the expectations of the measurement model?
3. How do the survey items and domains vary in their ability for respondents to positively agree with them?

Aim 2. The second aim of this study was to identify typologies of opportunity-to-learn using an unsupervised machine learning approach. The research questions that guided the second aim include:

1. Do meaningful opportunity-to-learn typologies exist based upon systematic differential item functioning (item-by-respondent) bias indices?
2. What are the predominant characteristics of the opportunity-to-learn typologies?

Aim 3. The third aim of this study was to build a Random Forest model in order to predict opportunity-to-learn classifications based upon systematic differential item functioning (respondent-by-item) bias indices. The research questions that guided the third aim include:

1. How accurately can a Random Forest model classify music programs into each of their respective opportunity-to-learn typologies?
2. What adjustments to the hyperparameters of the Random Forest model can be made to improve the model's error rate?
3. What are the most important survey items for predicting the opportunity-to-learn typologies of music programs?

Method

Apparatus. The NAfME Opportunity-to-Learn Standards are categorized into eight strands: (a) All Grades – All Content Areas, (b) PreK-2 General Music, (c) Grade 3-5 General Music, (d) Grade 6-8 (and all Secondary) General Music, (e) Ensembles (Elementary and Secondary Grades), (f) Composition/Theory, (g) Guitar/Keyboard/Harmonizing Instruments (Elementary and Secondary Grades), and (h) Technology. For this study, only secondary-level instrumental music programs were examined. Therefore, the criteria under the categories "All Content Areas" and "Ensembles (Elementary and Secondary Grades)" were used to generate survey item statements. First, content for survey items were extracted from the OTL Standards and formatted to produce short, concise, and useful item stems appropriate for a Likert-type rating scale. The syntax of the survey items were carefully extracted and written so as not to compromise the original syntax written in the standards themselves. Once compiled, survey items were reviewed and edited for clarity in word choice, relevance, redundancy, and directionality. The item stems were grouped into five broad domains based upon the OTL standards: (a) curriculum, (b) scheduling, (c) staffing, (d) materials and equipment, and (e) facilities. In order to appropriately and accurately develop a rating scale structure suitable to each OTL standard, each of the survey items was first identified as most appropriately eliciting either a dichotomous response (e.g., *yes/no*, *agree/disagree*, etc.) or polytomous, monotonically ordered response (e.g., s*trongly disagree, disagree, agree, strongly agree*; *never, rarely, occasionally, often, always*). Each item was mapped to an appropriate response anchor as outlined by Vagias (2006). The anchors used in this study included frequency (e.g., *never, rarely, occasionally, often always*) and level of agreement (e.g., *never, rarely, occasionally, sometimes, always*). The items were paired with an appropriate rating scale structure based upon the response anchors, with special attention given to an ordered, monotonic structure (e.g., word structures that suggest an increasing amount) across the rating scale structure. The rating scale structure did not include any "neutral" or "not applicable" categories in order to establish a forced choice response set and so not to interrupt the monotonic structure of the response anchors. The finalized item pool included a total of 112 items with rating scale categories ranging from 2-5 categories (See Appendix A).

Participants. Participants were recruited through the National Association for Music Education's (NAfME) Research Survey Assistance portal. The selection criteria included all 6-12, full-time, in-service music educators across the United States that teach secondary-level instrumental music. Prospective participants were contacted via two emails (one initial invitation email and one follow-up/reminder email). The survey remained accessible to prospective respondents for 20 days. Acceptance of informed consent was elicited by agreeing to participate in the study. Participants were ensured that their responses would be kept confidential throughout the entirety of the study. A total of 374 responses were collected, exceeding the necessary sample size for constructing measures using a measurement model with properties of invariant measurement (Linacre, 1994; Wright & Tennant, 1994) (See Figure 1).

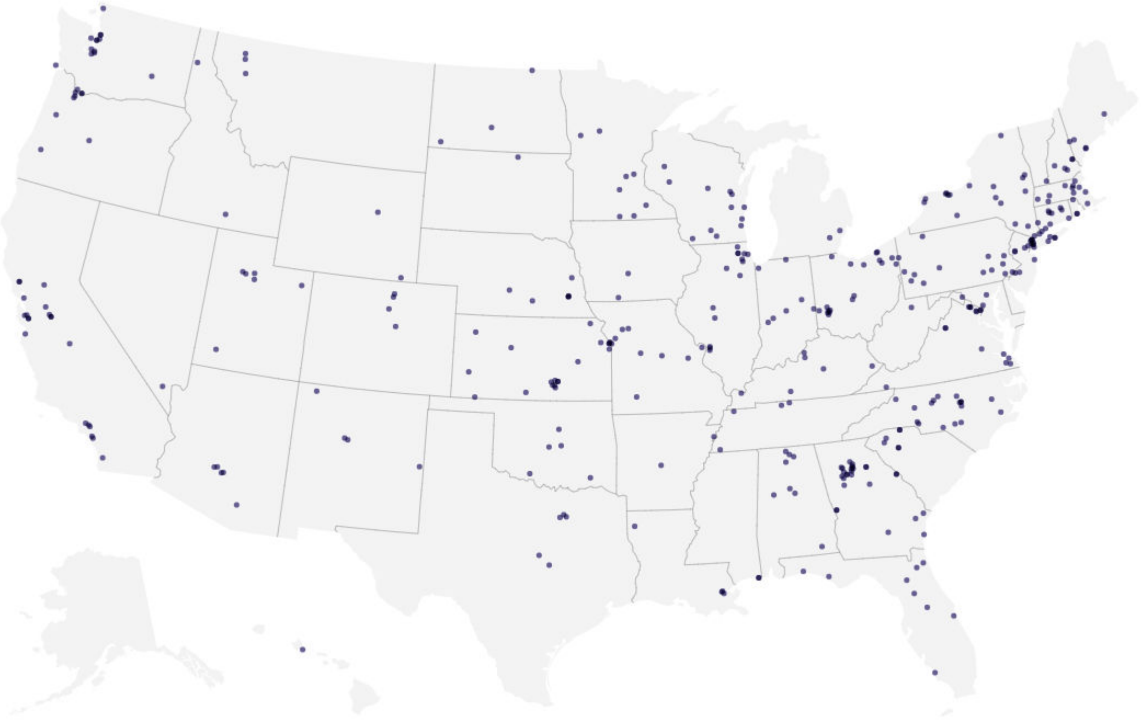


Figure 1. Participants Across the United States (*N* = 374)

Psychometric Considerations for Measuring Opportunity-to-Learn (Aim 1). Item Response Theory (IRT) is a broad, umbrella term describing a family of mathematical models that uses probabilistic distributions of raw score responses as a logistic function of person and item parameters in order to define unobservable, latent constructs (Wesolowski, 2019). Rasch measurement (Rasch, 1960/1980) is part of the IRT scaling tradition and is particularly effective for measuring latent constructs in the behavioral, social, and psychological sciences (Engelhard, 2013). The goal of the methodology is to produce meaningful and useful measures of latent constructs based upon a representative sample of items (Lord, 1980; Rasch, 1960/1980). In this study, the latent construct is conceptualized as "opportunity-to-learn" and the representative sample of items are the 112 survey items that operationally define the opportunity-to-learn construct. The FACETS computer program was used for all Rasch analyses (Linacre, 2014).

Psychometric Considerations for Identifying Differential Item Functioning (Aim 2). Differential item functioning (DIF) is defined as the examination of the conditional probabilities of a response to an item between respondents that have comparable locations on the latent variable (Wesolowski, 2015; Engelhard, 2013). In the FACETS computer program, an analysis of differential item functioning is referred to as a bias analysis. In this study, bias analyses were used to examine the interaction effects between the respondents (i.e., the music educator's responses to the survey items) and the probable response to the survey item based upon expectations of the model. According to Linacre (n.d.), bias estimates are important for several reasons: (a) diagnosing misfit- estimates of unexpected bias size help identify systematic misfit, (b) validity- bias in a survey item may not be detected by the usual summary fit statistics, and (c) effects- bias terms have a measure and a standard error and therefore can expressed in the same frame of reference as the survey item and/or respondent measures.

In the context of the research questions posed in this study, evidence of DIF not only provides validity evidence of the constructed OTL self-report rating scale, but also provides specific diagnostic evidence of respondents' interactions with the criteria set forth in the OTL standards beyond that of an initial analysis of model-data fit. The use of each item bias estimation for cluster detection (see below) is particularly helpful as it establishes a single interval-level (i.e., continuous) measure for the interaction between each respondent and each of the the individual survey items themselves.

Statistical Considerations for Cluster Detection (Aim 2). Machine learning is a branch of artificial intelligence that includes the automatic detection of complex data patterns by computing systems (Mitchell, 1997). Recently, with greater public access to big data, rapid advancements in computational performance, and the ability to glean volumes of data so massive that it surpasses the ability of humans to make sense of it, machine learning systems have provided a fruitful method for automating data-driven environments that learn from data through pattern recognition and, more importantly, use pattern recognition to learn from changes in data. Traditional applications of machine learning include demand forecasting, fraud detection, SPAM email detection, automated driving, targeted marketing, and preventative maintenance, for example (Mohammed, Khan, & Bashier, 2017). In education, applications of machine learning can provide a fruitful method for better understanding learning processes and providing a means toward understanding complex educational problems.

Unsupervised machine learning, a particular type of machine learning, is an investigatory method for detecting patterns from a dataset without reference to known (e.g., labeled) outcomes. In particular, one unsupervised machine learning methodology is to employ a particular set of algorithms to detect anomalous behaviors in data sets and map the data to spatially-coherent groups, or "clusters" (Celebi & Aydin, 2016). In order to address Aim 2, two unsupervised machine learning clustering algorithms were used to group respondents into similar classes based upon their respective bias indices gleaned from a item-by-respondent differential item functioning analyses. First, a hierarchical cluster analysis was used as an exploratory tool to investigate multiple cluster solutions for meaningfulness and substantive interpretability. Cluster solutions ranging between 2-10 were explored using Ward's Linkage method. Ward's method was chosen for its ability to partition interval-level variables into various cluster solutions by minimizing inter-class similarity and maximizing intra-class similarity. Squared Euclidian distances were used to compute the proximity distances between respondents. Second, after identifying a cluster solution with both a statistical and substantively meaningful interpretation, a non-hierarchical k-means cluster analysis was used to generate the most meaningful cluster solution from the hierarchical cluster analysis by specifically using the pre-specified cluster centroids. Using pre-specified cluster centroids allows for the iterative estimate of cluster assignments with greater case distinction while also providing an anchor for reproducible research. Both the hierarchical and non-hierarchical k-means cluster analyses were conducted in R statistics software (R Core Team, 2020).

Statistical Considerations for Classification Prediction (Aim 3). Supervised machine learning, another branch of machine learning, refers to a family of algorithms that train a statistical model of known input and output data with the intent of predicting uncertain, future outputs (Kotsiantis, 2007). Supervised machine learning algorithms analyze a certain percentage of a data set (i.e., training data) to produce an inferred function from another percentage of a data set (i.e., testing data) that can be used to map predicted labels (e.g., the opportunity-to-learn clusters) to new input data (e.g., a new survey response). The benefit of building a supervised machine learning model is the ability to provide new sets of categorical predictions using new and unfamiliar sets of input data without changes to the model specifications. In this case of this study, the goal is to create a model that can accurately and efficiently map an opportunity-to-learn classification to a newly completed survey response.

One popular supervised machine learning algorithm is Random Forests (Breiman, 2001). The Random Forests algorithm uses a single, binary classification, referred to as a decision tree, to make predictive classification decisions. A decision tree is a predictive input-output model that uses input variables (in this study, the 109 bias indices calculated in Aim 2) to make predictions about output variables (in this study, the 3 possible classifications identified in Aim 2). The trees, which are made up of nodes, represent the iterative possibility of all decisions. Parent nodes are created of various subsets of the input variables of the data set. These parent modes are iteratively partitioned into children nodes through a recursive process with the goal of decreasing impurity (i.e., a quantitative measure on which the optimal decision is based). Terminal nodes, or leaves, are labeled by an empirically calculated “best guess” of the output variable. A result of the process is an estimate of the error rate, or “out-of-bag (OOB) error.” The OOB error rate, interpreted more broadly as the model's overall predication accuracy, is one of the most important empirical indicators for validation of the Random Forest model. Another important key feature which makes Random Forests a popular choice for the analysis in this study is its ability to detect a hierarchy of variable importance for making classification decisions.

Results: Aim 1

Overall Calibration and Summary Statistics. Overall, the measure of opportunity-to-learn exhibited acceptable psychometric properties, indicating that the model estimates for respondents and survey items can be meaningfully interpreted. The model estimates for respondent measures explained 53.69% of the variance in their responses to the survey items, which exceeds the commonly used criterion of 20% for Rasch analyses of potentially multidimensional scales (Reckase, 1979). The

Table 1

Summary Statistics for the Rasch Measurement Model

	Facets	
	Respondents	**Survey Items**
Measure (Logits)		
Mean	-.01	0.00
SD	0.57	.92
N	374	104
Infit MSE		
Mean	1.03	1.00
SD	.30	.12
Std. Infit MSE		
Mean	.10	-.10
SD	1.90	.15
Outfit MSE		
Mean	1.00	1.00
SD	.24	.15
Std. Outfit MSE		
Mean	-.10	-.10
SD	1.60	2.10
Separation Statistics		
Reliability of Separation	.93	.99
Chi-Square	4657.00*	7968.20*
Degrees of Freedom	373	102

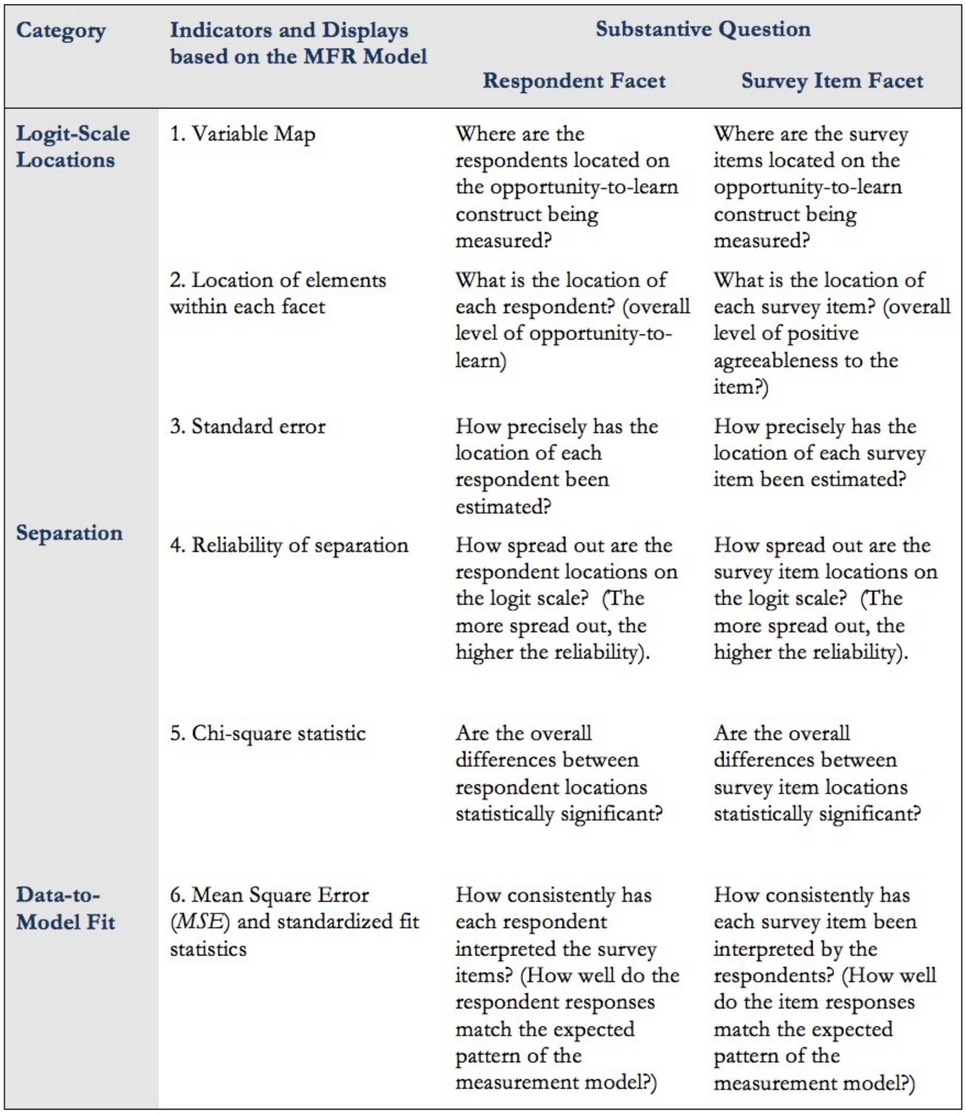
* *p* < .01

measure of opportunity-to-learn indicated statistically significant differences with high reliability of separation between respondents (χ^2(373) = 4657.00, *p* < .01, *Rel* = 0.93) and survey items (χ^2(102) = 7968.20, *p* < .01, *Rel* = 0.99). In the context of general linear modeling, this is comparable to demonstrating a significant main effect where respondents and survey items represent two independent variables. A wide range of respondent logit locations on the construct existed (-2.23-2.19) indicating that the survey items successfully identified respondents who exhibited low levels opportunity-to-learn, respondents who exhibited high levels opportunity-to-learn, and all continuous gradations in between. Furthermore, analysis of the residuals (differences between the responses we would observe if the model estimates were a perfect representation of respondents' response patterns and the actual observed responses) indicated that the model estimates were reasonable summaries of music educator responses to the opportunity-to-learn survey items, with average values of infit mean square error (MSE) and outfit MSE of around 1.00 (*M*Infit = 1.03, *SD* = 0.30; *M*Outfit = 1.00, *SD* = 0.24). The reliability of separation for respondent measures (similar to Cronbach's alpha coefficient) was 0.93, indicating that the survey items discriminated among respondents with different levels of opportunity-to-learn. Results of the Rasch analysis indicate that the measure performance has strong predictive validity as indicated by statistically significant results and high separation of reliability for the respondent facet and has strong construct validity as indicated by statistically significant results and high separation of reliability for the survey item facet. Parameter-level summary statistics for the model are found in Table 1 and the substantive interpretation for each statistic is found in Table 2.

Because the Rasch measurement model is unidimensional, it is possible to display the location estimates for each facet on a single, linear scale. The variable map is a useful method for visually displaying measures of respondents and survey items in terms of the latent variable. The usefulness of the variable map is a major factor in the adoption of Rasch modeling by many national and international surveys, including psychiatric outpatient surveys (Olsen, Garratt, Iversen, and Bjertnaes, 2010), US Household Food Security surveys (Kilanowski and Lin, 2012), and well-being of adoptive parents surveys (Furno, 2007), for example. Figure 2 is a variable map that is a graphical display of the latent variable (i.e., opportunity-to-learn) investigated in this study. Specifically, the map contains the calibrations of the two facets (i.e., variables) included in the model on the same linear scale (i.e., "a common "ruler"). In this

Table 2

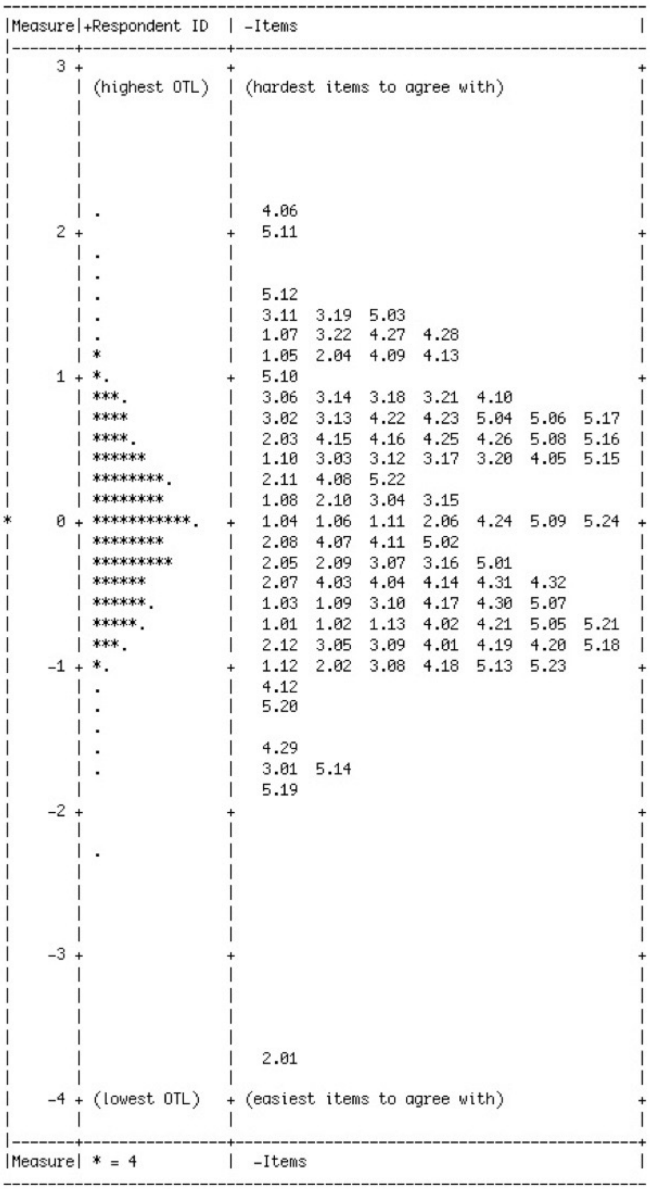
Substantive Interpretations of the Rasch Measurement Model Statistics



Category	Indicators and Displays based on the MFR Model	Substantive Question	
		Respondent Facet	Survey Item Facet
Logit-Scale Locations	1. Variable Map	Where are the respondents located on the opportunity-to-learn construct being measured?	Where are the survey items located on the opportunity-to-learn construct being measured?
	2. Location of elements within each facet	What is the location of each respondent? (overall level of opportunity-to-learn)	What is the location of each survey item? (overall level of positive agreeableness to the item?)
Separation	3. Standard error	How precisely has the location of each respondent been estimated?	How precisely has the location of each survey item been estimated?
	4. Reliability of separation	How spread out are the respondent locations on the logit scale? (The more spread out, the higher the reliability).	How spread out are the survey item locations on the logit scale? (The more spread out, the higher the reliability).
	5. Chi-square statistic	Are the overall differences between respondent locations statistically significant?	Are the overall differences between survey item locations statistically significant?
Data-to-Model Fit	6. Mean Square Error (*MSE*) and standardized fit statistics	How consistently has each respondent interpreted the survey items? (How well do the respondent responses match the expected pattern of the measurement model?)	How consistently has each survey item been interpreted by the respondents? (How well do the item responses match the expected pattern of the measurement model?)

study, the two facets were respondents and survey items. Column 1 includes the units of the logit scale whereby all facets can be calibrated and compared. Column 2 includes the spread of respondent calibrations (i.e., levels of opportunity-to-learn), where each asterisk represents four respondents. The top of the column represents the highest levels of opportunity-to-learn and the bottom of the column represents the lowest levels of opportunity-to-learn. Column 3 represents the spread of the survey item calibrations. The top of the column represents the survey items that were the hardest for respondents to agree with and the bottom of the column represents the survey items that were the easiest for the respondents to agree with.

Calibration of Respondent Facet. Appendix B provides the detailed calibration information for each of the respondents. Acting as the object of measurement, the respondent facet was allowed to float (i.e., the mean was non-centered). The mean logit score for the respondent facet was -.01 logits. The range of respondent measures was from 2.19 logits for the highest scoring respondent (respondent 50) to -2.23 logits for the lowest scoring respondent (respondent 87). The highest scoring respondent is interpreted as having the highest opportunity-to-learn measure for their respective music program and the lowest scoring respondent is interpreted as having the lowest opportunity-to-learn measure for their respective music program. Based upon acceptable fit criteria of 0.60-1.40 for survey data as indicated by Wright and Linacre (1994), 41 of the 374 respondents (11.00%) did not reasonably fit the model. These respondents were removed from the data set for subsequent analyses.



```
----------------------------------------------------------------------------
|Measure|+Respondent ID  | -Items                                          |
|-------+----------------+-------------------------------------------------
|     3 +                +                                                 +
|       | (highest OTL)  | (hardest items to agree with)                   |
|       |                |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|       | .              |  4.06                                           |
|     2 +                +  5.11                                           +
|       | .              |                                                 |
|       | .              |                                                 |
|       | .              |  5.12                                           |
|       | .              |  3.11  3.19  5.03                               |
|       | .              |  1.07  3.22  4.27  4.28                         |
|       | *              |  1.05  2.04  4.09  4.13                         |
|     1 + *.             +  5.10                                           +
|       | ***.           |  3.06  3.14  3.18  3.21  4.10                   |
|       | ****           |  3.02  3.13  4.22  4.23  5.04  5.06  5.17       |
|       | ****.          |  2.03  4.15  4.16  4.25  4.26  5.08  5.16       |
|       | ******         |  1.10  3.03  3.12  3.17  3.20  4.05  5.15       |
|       | ********.      |  2.11  4.08  5.22                               |
|       | ********       |  1.08  2.10  3.04  3.15                         |
*     0 + ***********.   +  1.04  1.06  1.11  2.06  4.24  5.09  5.24       +
|       | ********       |  2.08  4.07  4.11  5.02                         |
|       | *********      |  2.05  2.09  3.07  3.16  5.01                   |
|       | ******         |  2.07  4.03  4.04  4.14  4.31  4.32             |
|       | ******.        |  1.03  1.09  3.10  4.17  4.30  5.07             |
|       | *****.         |  1.01  1.02  1.13  4.02  4.21  5.05  5.21       |
|       | ***.           |  2.12  3.05  3.09  4.01  4.19  4.20  5.18       |
|    -1 + *.             +  1.12  2.02  3.08  4.18  5.13  5.23             +
|       | .              |  4.12                                           |
|       | .              |  5.20                                           |
|       | .              |                                                 |
|       | .              |  4.29                                           |
|       | .              |  3.01  5.14                                     |
|       |                |  5.19                                           |
|    -2 +                +                                                 +
|       |                |                                                 |
|       | .              |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|    -3 +                +                                                 +
|       |                |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|       |                |                                                 |
|       |                |  2.01                                           |
|       |                |                                                 |
|    -4 + (lowest OTL)   + (easiest items to agree with)                   +
|       |                |                                                 |
|-------+----------------+-------------------------------------------------+
|Measure| * = 4          |  -Items                                         |
----------------------------------------------------------------------------
```

Figure 2. Variable Map of the Opportunity-to-Learn Construct

Calibration of Domain and Survey Item Facets. The survey items were grouped into five domains: (a) curriculum (*n* = 13); (b) scheduling (*n* = 12); (c) staffing (*n* = 23); (d) materials and equipment (*n* = 34); and (e) facilities (*n* = 24). The rank ordering from the hardest domain for respondents to positively agree with (i.e., lowest logit measure) to the easiest domain for respondents to positively agree with (i.e., highest logit measure) were the: (a) staffing domain (*M* = .24 logits, *n* = 22, χ^2(21) = 2074, *Rel* = .99, *p* < .01); (b) materials and equipment domain (*M* = 0.03 logits, *n* = 32, χ^2(31) = 2761.60, *Rel* = .99, *p* < .01); (c) facilities domain (*M* = 0.00 logits, *n* = 24, χ^2(23) = 1461.60, *Rel* = .99, *p* < .01); (d) curriculum domain (*M* = -0.11 logits, *n* = 13, χ^2(12) = 534.50 , *Rel* = .98, *p* <. 01); and (e) scheduling domain (*M* = -0.39 logits, *n* = 12, χ^2(11) = 534.50, *Rel* = .99, *p* < .01). All domains demonstrated reasonable fit to the model, suggesting that each domain, as a group of items, can be meaningfully interpreted (See Table 3).

A Pearson's *r* product-moment correlation was used to investigate the relationship between each of the domains of the opportunity-to-learn scale. Results indicated statistically significant correlations

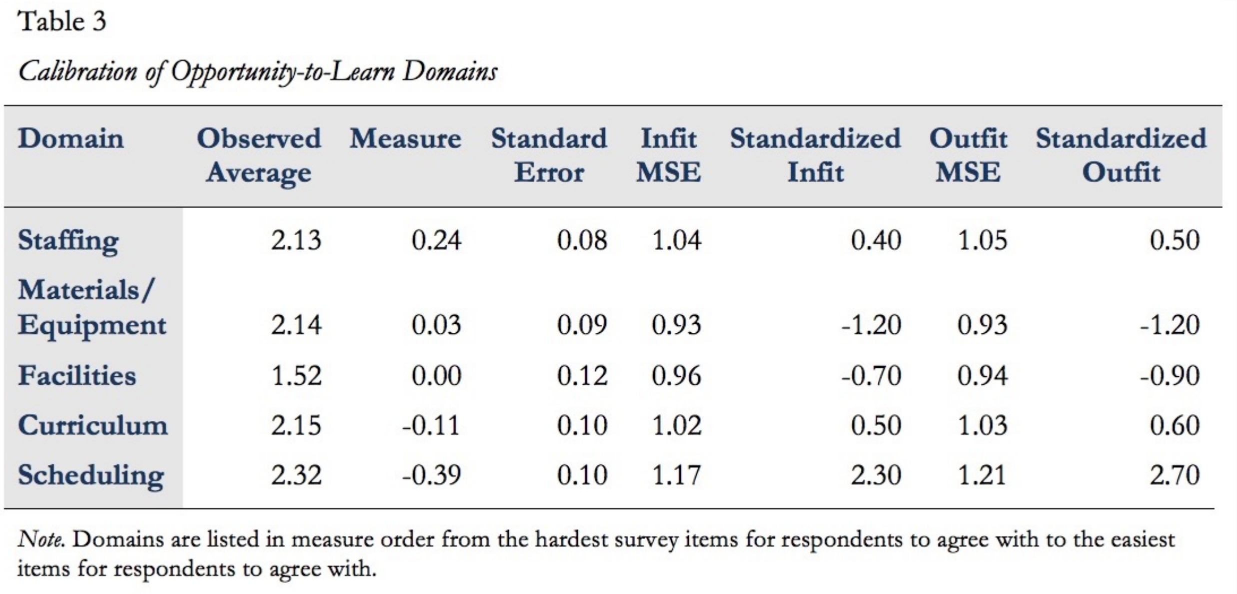


Table 3

Calibration of Opportunity-to-Learn Domains

Domain	Observed Average	Measure	Standard Error	Infit MSE	Standardized Infit	Outfit MSE	Standardized Outfit
Staffing	2.13	0.24	0.08	1.04	0.40	1.05	0.50
Materials/ Equipment	2.14	0.03	0.09	0.93	-1.20	0.93	-1.20
Facilities	1.52	0.00	0.12	0.96	-0.70	0.94	-0.90
Curriculum	2.15	-0.11	0.10	1.02	0.50	1.03	0.60
Scheduling	2.32	-0.39	0.10	1.17	2.30	1.21	2.70

Note. Domains are listed in measure order from the hardest survey items for respondents to agree with to the easiest items for respondents to agree with.

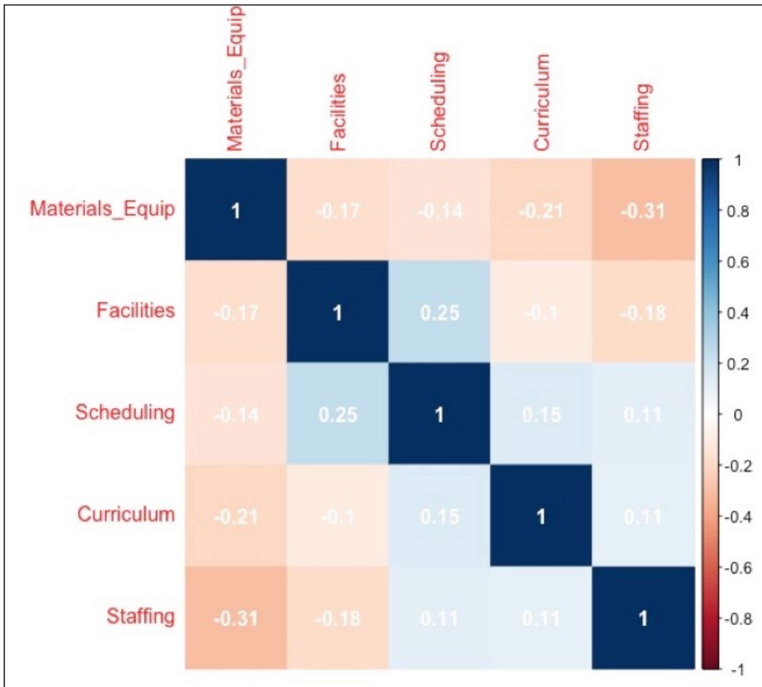


Figure 3. Correlation Matrix of Opportunity-to-Learn Domains

(*p* < .05) between all domains. Correlations and confidence intervals can be found in Figures 3 and 4, respectively.

Appendix B provides the calibration information for each of the survey items. The mean was centered at 0.00 logits in order to provide a better frame of reference of the overall interpretation of the respondent facet. The range of survey item measures was from 2.14 logits for the survey item that was, overall, hardest for respondents to positively agree with (i.e., highest logit measure) (Item 4.06: *The school program has a written depreciation and replacement plan for all instruments and equipment, specifically describing under what conditions instruments should be retired and replaced*) to -3.74 logits for the survey item that was, overall, easiest for respondents to positively agree with (i.e., lowest logit measure) (Item 2.01: *Every performing group presents a series of performances for parents, peers, or the community*).

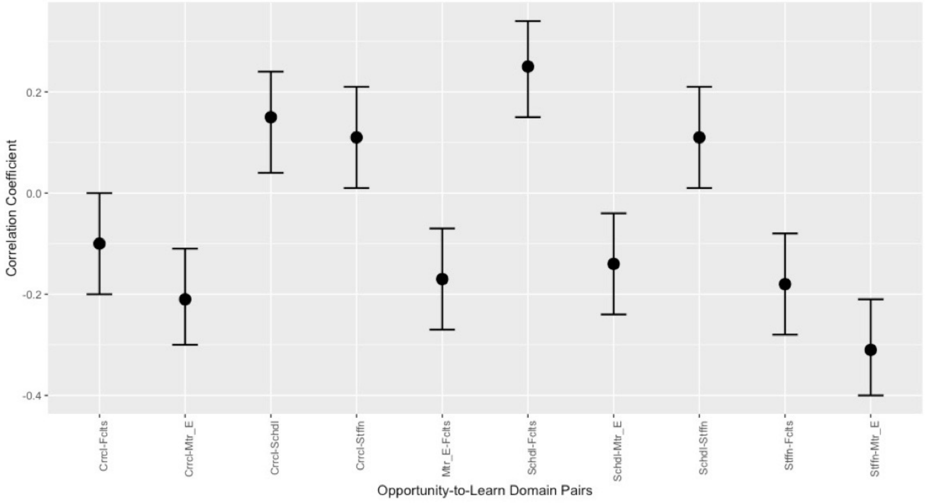


Figure 4. Correlations and Confidence Intervals for Opportunity-to-Learn Domain Pairs. Crrcl = Curriculum, Fclts = Facilities, Mtr_E = Materials and Equipment, Scdul = Scheduling, Stffn = Staffing

Results: Aim 2

In order to calculate bias indices for each participant's response to each survey item, a differential item functioning (DIF) analysis was conducted by crossing the respondent and survey item facets. A total of 36,297 interactions (333 remaining respondents after removing misfits multiplied by 109 remaining items after removing misfits) were extrapolated from the analysis. A data frame was developed that contained 333 rows of respondents and 109 columns of bias index variables. In order to prepare the bias indices for cluster analysis, a [0,1] min-max scaling was conducted on each of the variables. In cluster analyses, clusters are defined by the distance between points in mathematical spaces, or dimensions. Therefore, scaling the range of the indices to a [0,1] range allows each variable to be examined using approximately proportionate distances, thereby contributing equal weight to the analyses (Ioffe and Szegedy, 2015).

Two types of cluster analyses were used to explore the meaningfulness of possible groupings of respondents based upon the differential item functioning measures (i.e., bias indices) from the respondent-by-item interaction. A hierarchical cluster analysis was first used in order to explore multiple cluster solutions for both statistical and substantive meaningfulness. In order to partition the survey items based upon the respondent-by-item bias indices, Ward's linkage agglomerative clustering method was used. Ward's method was specifically used in order to determine the degree of acceptability in which clusters are linked together by maximizing intra-class similarity and minimizing inter-class similarity. Additionally, it is regarded as the most efficient linkage procedure through the minimization of within-cluster sums of squares over all partitions available (Ward, 1963). Squared Euclidian distances were selected as a method for computing the proximity between respondents. The advantage of using this method is that possible outliers do not affect distances between respondents. Additionally, it places progressively greater weights on respondents further apart (Romesburg, 1984). A principal components analysis was conducted in order to examine which survey items contribute most to the two-dimensional space based upon their squared Euclidian distances. The ten highest-weighted survey items contributing to Dimension-1 and the ten highest-weighted survey items contributing to Dimension-2 can be found in Figure 5. Based on the interpretation that Dimension 1 and Dimension

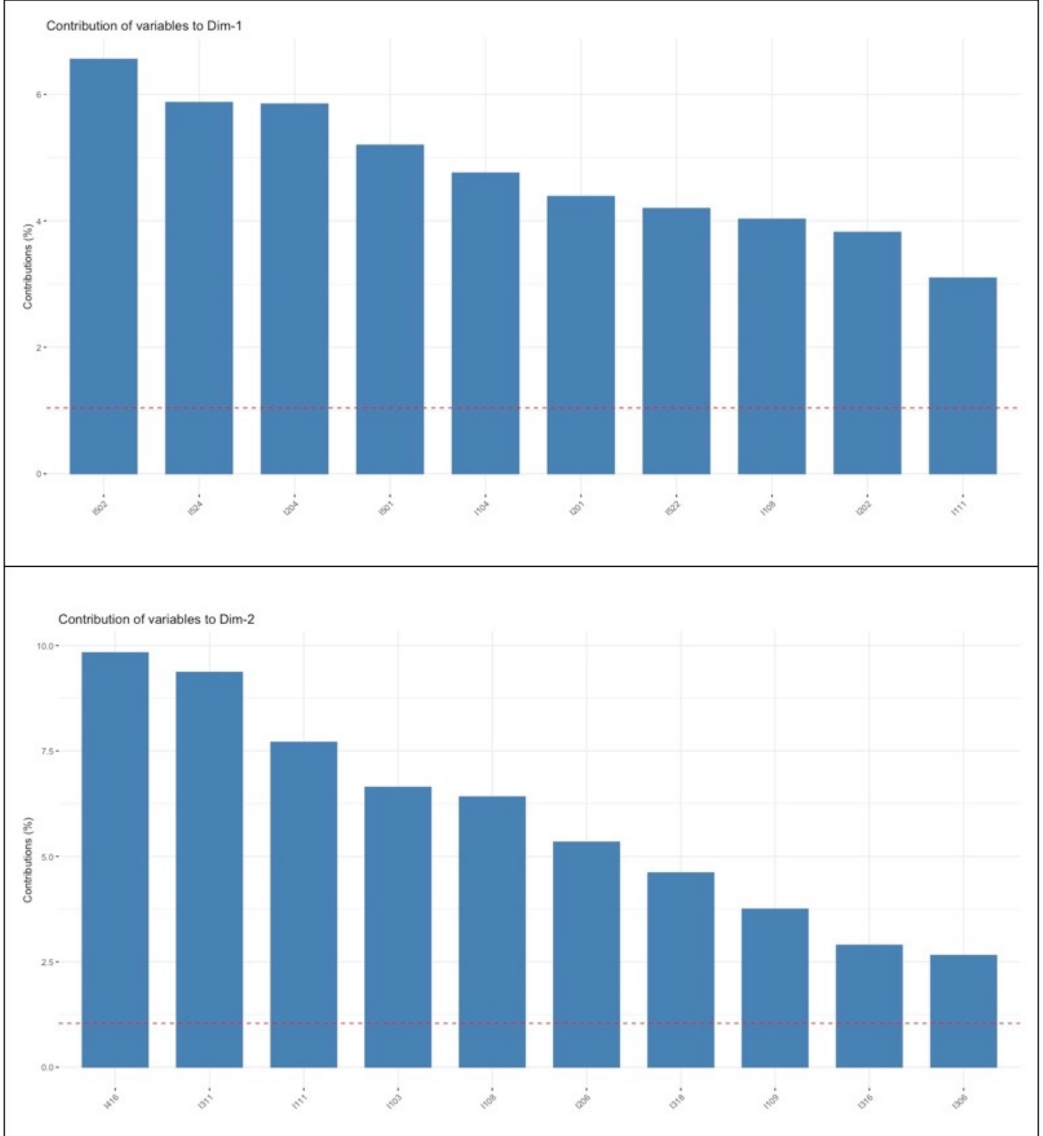


Figure 5. Contribution of Survey Items to Dimension 1 and Dimension 2 from the Principal Components Analysis

2 as opposite in nature, we see that Dimension 1 is grounded in a focus of facilities. In particular, the survey items from Dimension 1 reflect adequacy and quality of performance spaces (survey items 5.02, 5.24, 5.01, 5.22) and scheduling of performance series in those spaces (survey item 2.02), for example. Oppositely, we see that Dimension 2 is grounded in a focus of student-centered opportunities and student learning outcomes. In particular, the survey items from Dimension 2 reflect special opportunities for students to engage in musical experiences through diversity in curricula (survey items 1.11, 1.03, 1.09) and scheduling priorities (survey item 2.06), as well as music teacher evaluations based upon music-specific learning outcomes (survey items 3.18 and 3.16), for example. The full contribution of items to the two-dimensional space is provided in Figure 6.

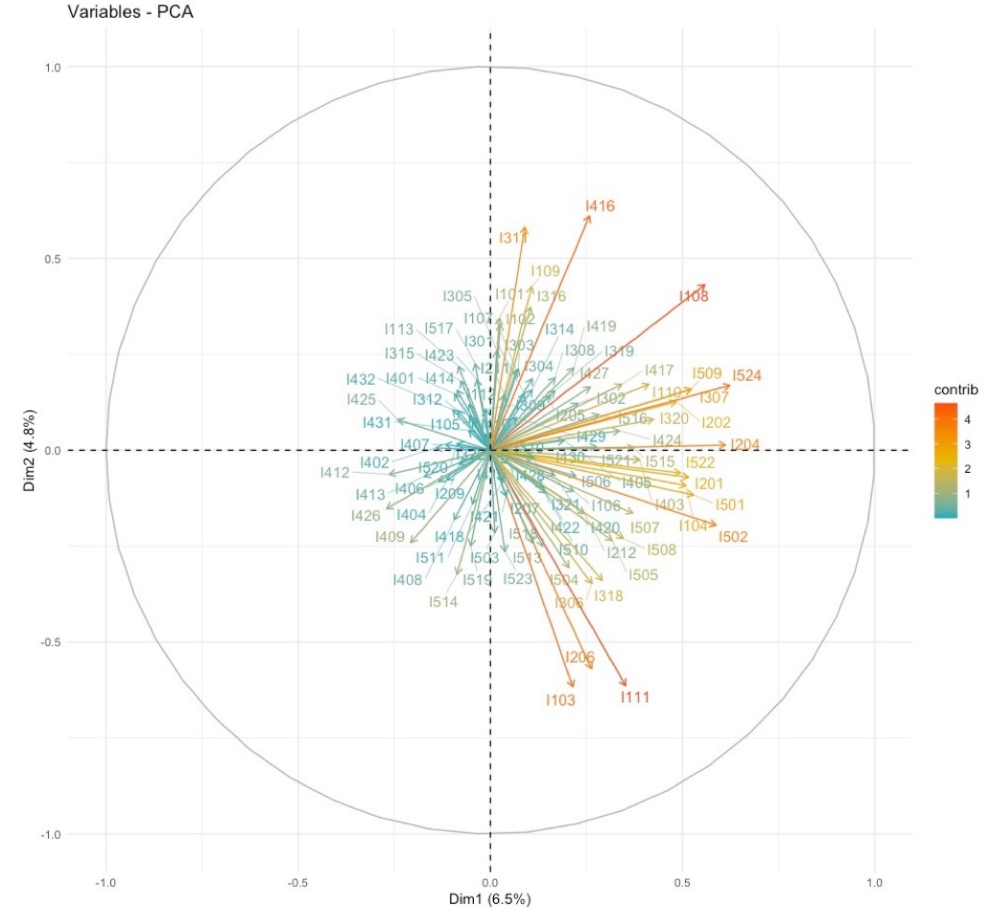


Figure 6. Contribution of all Survey Items to the Dimensional Space

In order to identify the most meaningful cluster solution, clusters ranging from 2-10 were examined for reasonably discernible and substantive trends. Additionally, Mardia, Kent, and Bobby's (1979) "rule of thumb," Thorndike's (1953) elbow method, and a silhouette analysis (Rousseau, 1987) were considered for a quantitative interpretation of potential cluster solutions. Based upon these considerations, a three-cluster solution was selected as the most substantively interpretative.

After selecting a three-cluster solution as the most meaningful interpretation, a follow-up non-hierarchical k-means cluster analysis was used to generate a three-cluster solution with optimal case distinctions. A k-means clustering technique was appropriate in order to iteratively estimate cluster means based upon smallest distances to the cluster mean. Specifically, the three-cluster solution cluster centroids from the hierarchical cluster analysis were used as seeds (i.e., cluster centroids as anchor means) to pre-specify the threshold distances between respondents (Figure 7).

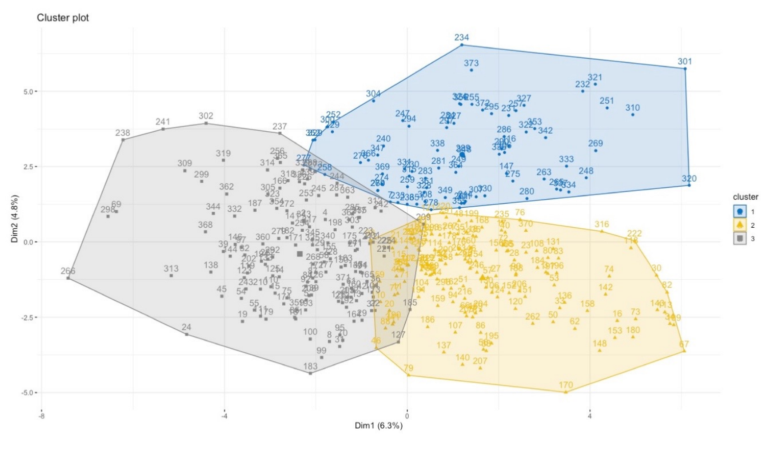


Figure 7. Three-cluster Solution for Possible Opportunity-to-Learn Typologies

Table 4 provides the finalized cluster centroids for each survey item by cluster assignment. Note that for each of the three cluster values mapped to an item (e.g., for item 1.01, cluster 1 = 0.50, cluster 2 = -0.62, cluster 3 = -0.58) one cluster centroid value has a distinctly different positive or negative value as indicated by shading (in the case of item 1.01, cluster 1 has a positive value of 0.50). Interpreting 0.00 as the mean/median for the values of each cluster dimensions provides an anchor for interpretation of the centroid values. Centroid values below 0.00 can be interpreted as having a lower weighted value than other clusters. Conversely, centroid values above 0.00 can be interpreted as having a higher weighted value than other clusters. Investigating values on the opposite side of 0.00 from the other clusters is significant in differentiating clusters by each survey item's weighted value. The data gleaned from Table 4 played the primary role in making substantive interpretations about the clusters based upon item behavior. The finalized cluster solution explained 31.47% of the variance attributed to the bias indices.

Cluster 1. A total of 34.53% (*n* = 115) of respondents comprised cluster 1. According to the differentiation of cluster centroids as demonstrated in Table 4, cluster 1 is distinguished by the following characteristics:

- Strong depth and breadth of curriculum (items 1.01, 1.02, 1.05, 1.06, 1.07, 2.05, 3.16);
- Strong instructional considerations for students with disabilities (items 3.03, 3.04 3.11, 3.12);
- Focus on professional development (items 3.05, 3.07, 3.09, 3.10) and music-based teacher evaluation (3.17, 3.19);
- Positive administrative consideration for scheduling music classes (items 2.06, 2.07); and
- Insufficient amount of equipment (item 4.12), materials (items 4.22), technology (items 4.12, 4.26, 4.27, 4.28, 4.30, 4.32, 4.33, 4.34) and lack in facilities (items 5.01-5.09, 5.15-5.19, 5.22-5.23).

Table 4

Finalized Cluster Centroids by Survey Item

Cluster	Centroids by Survey Item								
	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09
1	0.50	0.32	0.20	0.12	0.14	0.20	0.32	0.26	0.20
2	-0.62	-0.26	0.12	0.15	-0.09	-0.10	-0.17	-0.42	-0.37
3	-0.58	-0.08	-0.35	-0.29	-0.06	-0.12	-0.17	0.16	0.18
	1.10	1.11	1.12	1.13	2.01	2.02	2.05	2.06	2.07
1	0.26	-0.04	0.13	0.30	-0.56	0.58	0.45	-0.52	-0.57
2	-0.49	0.14	0.17	0.07	-0.67	-0.68	-0.59	0.41	0.59
3	0.23	-0.10	-0.33	-0.41	0.54	-0.59	-0.50	0.31	0.57
	2.08	2.09	2.11	2.12	3.01	3.02	3.03	3.04	3.05
1	-0.15	0.99	0.39	0.59	0.60	0.49	0.46	0.49	0.47
2	-0.09	-0.98	-0.45	-0.63	-0.68	-0.59	-0.51	-0.54	-0.58
3	0.27	0.98	0.46	0.51	0.62	0.53	-0.49	0.49	-0.54
	3.06	3.07	3.08	3.09	3.10	3.11	3.12	3.13	3.14
1	0.55	0.57	0.73	0.55	0.49	0.22	0.50	0.25	0.45
2	0.51	-0.67	-0.79	-0.57	-0.51	-0.40	-0.50	-0.57	-0.47
3	-0.31	-0.56	0.75	-0.55	-0.50	-0.46	-0.48	0.33	0.43
	3.15	3.16	3.17	3.18	3.19	3.20	3.21	3.22	3.23
1	0.49	0.42	0.32	0.58	0.35	-0.52	0.46	0.12	-0.01
2	-0.49	-0.59	-0.29	0.52	-0.39	0.58	0.45	0.09	0.32
3	0.53	-0.56	-0.05	-0.37	-0.38	-0.52	-0.43	-0.07	-0.32
	4.01	4.02	4.03	4.04	4.05	4.06	4.07	4.08	4.09
1	0.44	-0.55	-0.43	0.60	-0.40	-0.12	0.50	-0.50	-0.35
2	0.48	0.53	0.54	0.59	0.42	0.02	0.54	0.45	0.21
3	-0.52	-0.55	-0.43	-0.62	-0.42	-0.08	-0.54	-0.46	-0.30
	4.10	4.11	4.12	4.13	4.14	4.15	4.16	4.17	4.18
1	0.44	-0.39	-0.72	-0.36	0.57	0.46	0.25	-0.54	-0.55
2	0.46	0.61	0.65	0.24	0.57	0.44	-0.50	-0.64	0.57
3	-0.45	-0.22	0.76	-0.35	-0.59	-0.48	0.51	0.57	-0.57
	4.19	4.20	4.21	4.22	4.23	4.24	4.25	4.26	4.27
1	0.48	-0.53	-0.70	-0.53	0.50	-0.62	-0.46	-0.59	-0.67
2	0.59	-0.58	0.71	0.54	-0.53	0.66	0.40	0.43	0.77
3	-0.53	0.52	-0.69	0.54	-0.55	-0.62	-0.51	0.53	0.68
	4.28	4.29	4.30	4.31	4.32	4.33	4.34	5.01	5.02
1	-0.51	-0.48	-0.53	0.45	-0.51	-0.08	-0.14	-0.45	-0.49
2	0.53	0.53	0.52	-0.45	0.47	0.06	0.12	0.57	0.59
3	0.51	0.48	0.50	-0.47	0.52	0.02	0.03	0.42	0.43
	5.03	5.04	5.05	5.06	5.07	5.08	5.09	5.10	5.11
1	-0.13	-0.33	-0.62	-0.51	-0.48	-0.31	-0.57	-0.31	-0.12
2	0.09	0.29	0.66	0.53	0.53	0.32	0.67	-0.30	-0.04
3	0.07	0.27	0.57	0.52	0.43	0.27	0.56	0.31	0.07
	5.12	5.13	5.14	5.15	5.16	5.17	5.18	5.19	5.20
1	-0.24	-0.71	-0.96	-0.53	-0.38	-0.47	-0.72	-0.70	-0.70
2	0.34	-0.71	-0.88	0.58	0.44	0.53	0.71	0.69	-0.69
3	-0.10	0.64	0.91	0.53	0.41	0.61	0.66	0.73	0.73
	5.21	5.22	5.23	5.24					
1	-0.69	-0.51	-0.74	0.52					
2	-0.75	0.56	0.72	-0.68					
3	0.67	0.49	0.63	0.54					

Note. Lighter shaded areas indicate differentiation from other clusters above 0.00. Darker shaded areas indicate differentiation from other clusters below 0.00.

Cluster 2. A total of 32.43% (*n* = 108) of respondents comprised cluster 2. According to the differentiation of cluster centroids as demonstrated in Table 4, cluster 2 is distinguished by the following characteristics:

- Supported by materials and equipment (items 4.02, 4.03, 4.05, 4.06, 4.08, 4.09, 4.11, 4.13, 4.18, 4.21, 4.24, 4.25, 5.12);
- Lack of staffing, staffing qualifications, and staff development (items 3.01, 3.02, 3.08, 3.15);
- Scheduling and special education access concerns (items 1.09, 1.10, 2.09, 2.11, 2.12, 5.24); and
- Teaching load concerns (items 3.31, 3.14).

Cluster 3. A total of 33.03% (*n* = 110) of respondents comprised cluster 3. According to the differentiation of cluster centroids as demonstrated in Table 4, cluster 3 is distinguished by the following characteristics:

- Provided access to musical literature (items 4.17, 4.20);
- Lack of funding for purchasing and maintaining instruments (items 4.01, 4.04, 4.07, 4.10, 4.15);
- Sufficient facilities (items 5.10, 5.11, 5.13, 5.14, 5.20, 5.21);
- Lacks depth/breadth in curriculum (items 1.03, 1.04, 1.12, 1.13); and
- Staffing concerns (items 3.21, 3.22).

Results: Aim 3

Dataset and Model Preparation. The cluster assignments for each of the respondents (*n* = 333) were attached to the bias analysis data frame for the classification prediction. Note that in unsupervised machine learning methods (Aim 2), the grouping assigned to a respondent is referred to as a cluster. In supervised machine learning methods, because the cluster is already established and assigned to respondents a priori, the grouping is referred to as a classification. In the context of standard machine learning vernacular (Google Developers, 2020), the respondent-by-item bias index variables (*N* = 109) and opportunity-to-learn classification variable (*N* = 1) are individually referred to as feature vectors and collectively referred to as a feature set. In the case of the Random Forest model built in this study, the 109 continuous feature vectors (respondent-by-item bias indices) were used to make predictions based upon the 1 discrete feature vector (opportunity-to-learn classifications). The opportunity-to-learn classifications were labeled as Cluster 1 (*n* = 115), Cluster 2 (*n* = 108), and Cluster 3 (*n* = 110).

In order to avoid overfitting the model, the data frame was randomly subset into a 70% training data set (*n* = 233 respondents) and 30% testing data set (*n* = 100 respondents) (Gholamy, Kreinovich, Kosheleva, 2018). Based upon the recommendation of Breiman (2001), three hyperparameters were included in the model: (a) the initial number of trees to grow (i.e., ntree) specified in the initial model was set to 500 trees, (b) the number of random variables initially sampled at each split (i.e., mtry) was set to 10 variables, which by default is the rounded square root of the number of input variables included in the

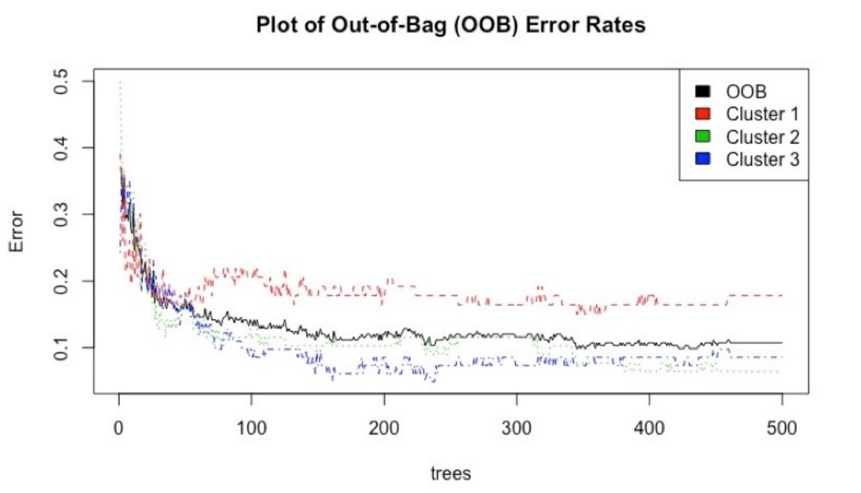


Figure 8. Plot of the Out-of-Bag (OOB) Error Rates for the Overall Random Forest Model and each of the Three Classifications

data set (*n* = 109), and (c) the minimum node size of each tree (i.e., node size) was set to a default of one node.

Default Random Forest Model Training. The initial model specifications consisted of three hyperparameters set to the default settings as specified by Breiman (2001): ntree (*n* = 500), mtry(*n* = 10), and and node size (*n* = 1). Based upon the default hyperparameters, the out-of-bag (OOB) error rate was 13.30%. Figure 8 provides the OOB error plots for each opportunity-to-learn classification as well as the overall model across the default ntree (*n* = 500) specification. The initial model accuracy, interpreted as the ratio of correctly predicted classifications to all possible classifications, was 87.00% (95% CI[0.81, 0.94]). Overall, 26 of the 233 total cases were misclassified in the initial model specification (See Table 5).

Table 5

Confusion Table for Initial Model Specification and Tuned Hyperparameter Model

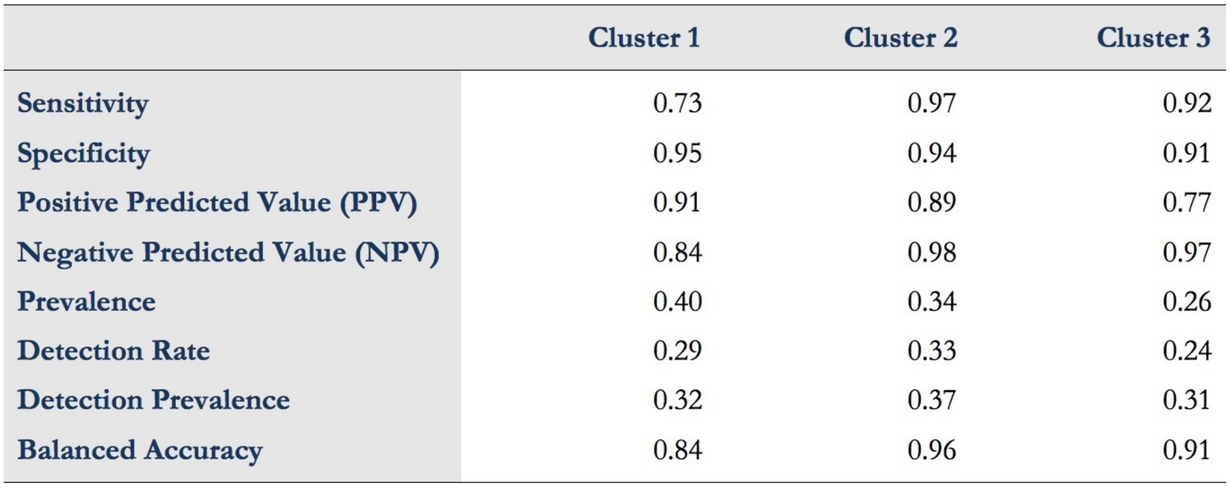
	Actual Classifications			
Predicted Classifications	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Class Error**
Initial Model Specification				
Cluster 1	61	10	2	0.16
Cluster 2	6	71	1	0.09
Cluster 3	6	1	75	0.09
Tuned Hyperparameter Model				
Cluster 1	60	9	4	0.17
Cluster 2	3	73	2	0.06
Cluster 3	5	2	75	0.09

In Table 5, the columns refer to the actual classifications and the rows refer to the predicted classifications. The diagonal (from top left to bottom right) indicates the true positives and true negatives. The top right of the diagonal indicates the false positive classifications, and the bottom left of the diagonal indicates the false negative classifications.

Table 6 provides the summary statistics for the initial model specification. Sensitivity is the proportion of relevant results out of the number of samples that were actually relevant. Specificity the proportion of true negatives that are correctly identified by the model. Positive predictive value (PPV) is the proportion of OTL class matching the corresponding OTL class correctly classified. Negative predictive value (NPV) is the proportion of OTL classes not matching the corresponding OTL classes correctly classified. Prevalence is interpreted as how often each category occurs in the sample. The detection rate is the rate of true events correctly predicted by the events. Detection prevalence is the commonness of predicted events. Balanced accuracy is the average accuracy obtained for all three OTL classes. Overall, the model demonstrates strong predictive accuracy, relatively low error, and therefore, a strong case for validity.

Table 6

Summary Statistics for the Initial Random Forest Model by Opportunity-to-Learn Classification



	Cluster 1	Cluster 2	Cluster 3
Sensitivity	0.73	0.97	0.92
Specificity	0.95	0.94	0.91
Positive Predicted Value (PPV)	0.91	0.89	0.77
Negative Predicted Value (NPV)	0.84	0.98	0.97
Prevalence	0.40	0.34	0.26
Detection Rate	0.29	0.33	0.24
Detection Prevalence	0.32	0.37	0.31
Balanced Accuracy	0.84	0.96	0.91

Tuning Hyperparameters and Optimizing Model. When creating a Random Forest Model, the initial specifications using default hyperparamters is an important first step in exploring the ability of the model to make intelligent, predictive predictions. Based on the initial specifications, an OOB error rate of 13.30% demonstrates a strong capability of the model. The three default hyperparameters used in this model were (a) ntree; *n* = 500 (the number of trees in the forest) (b) mtry; *n* = 10 (the number of variables randomly sampled as candidates for each split), and (c) node size; *n* = 1 (the minimum number of samples on the terminal nodes). Conceptually, tuning the hyperparameters to the model is similar to turning a combination of dials on a AM/FM radio in order to find the clearest sound. Here, we can test all combinations of hyperparameters to find the best combination with the lowest OOB error rate.

In order to iterate through the various combinations of the three hyperparameters, the tuneRF() function with the doBest parameter set to true (from the {randomForest} R package) was used. Tuning the model included two important steps. First, an iteration through all mtry values was conducted to identify the value that best minimizes OOB Error. Second, a list of all possible value combinations for mtry, nodesize and sample size (via tree depth) was created and used to create a specialized data frame referred to as a hyper-grid. The process of conducting a manual grid search included establishing a data frame of all possible combinations of values for mtry, node size, sample size, and training iterative models consisting of each combination in order to identify the optimal set of hyperparameters based on minimal OOB error. After establishing the manual grid search, the results indicated an optimized model with an mtry value of 8 and node size value of 5. The results of the tuning process suggested that the optimized model, using the set of tuned hyperparameter specifications of ntree = 40, mtry = 8 and node size = 5, demonstrated an OOB error rate of 10.73%. The newly tuned model improved performance by an accuracy rate of 2.57% (See Table 6, Tuned Hyperparameter Model).

Importance of Variables and the Mean Decrease GINI Value. The popularity of the Random Forest model is due to its ability to identify the feature vectors (i.e., variables) that contribute the most amount of weight to the classification process. An innate artifact of the decision tree algorithm process is a feature selection process, which highlights the most efficient and relevant features for maximizing correctly predicted partitioning of the data. The Gini variable importance measure (i.e., mean

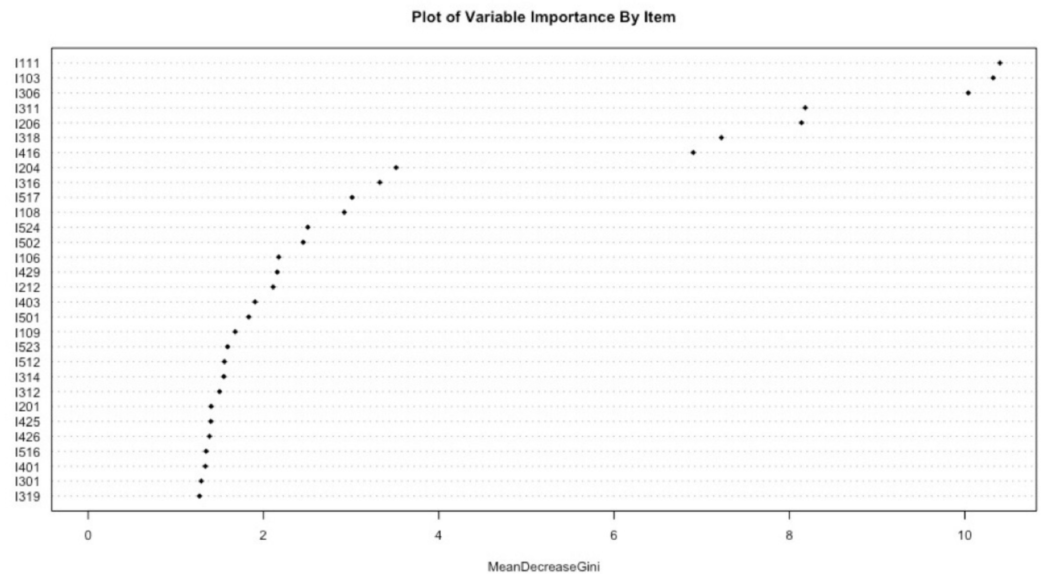


Figure 9. Plot of Survey Items by Importance in the Classification Process

decrease Gini) and root square mean square error (RMSE) are two important metrics that provide empirical evidence of these features' importance. Figure 9 is a plot of variable importance and Figure 10 is a plot of the recursive feature vector selection for the 30 most influential feature vectors used in the Random Forest model. Based on the interpretation of the plot of variable importance (Figure 9), there is a clear partitioning between the first seven items in terms of importance to identifying the unique nature of each of the three opportunity-to-learn classifications: Item 1.11 (*Gifted and talented are students provided with access to special experiences according to their abilities and interests*), Item 1.03 (*Your school offers at least one alternative performing organization or emerging ensemble (e.g. jazz band) for every 450 students in the school population*), Item 3.06 (*Each school district or school provides a regular program of in-service education arranged by the district or school each year for every music educator*), Item 3.11 (*Every music educator working with special education students has received sufficient in-service training in special education*), Item 2.06 (*Effort is made to avoid scheduling single section music classes against single section classes in other subjects*), Item 3.18 (*If student performance data are considered in teacher evaluation, data must involve music outcomes*), and Item 4.16 (*There is a folder of original music for each stand of no more than two instrumentalists*). The recursive feature selection plot (Figure 10) suggests that the first five of the seven items have the most predictive efficiency.

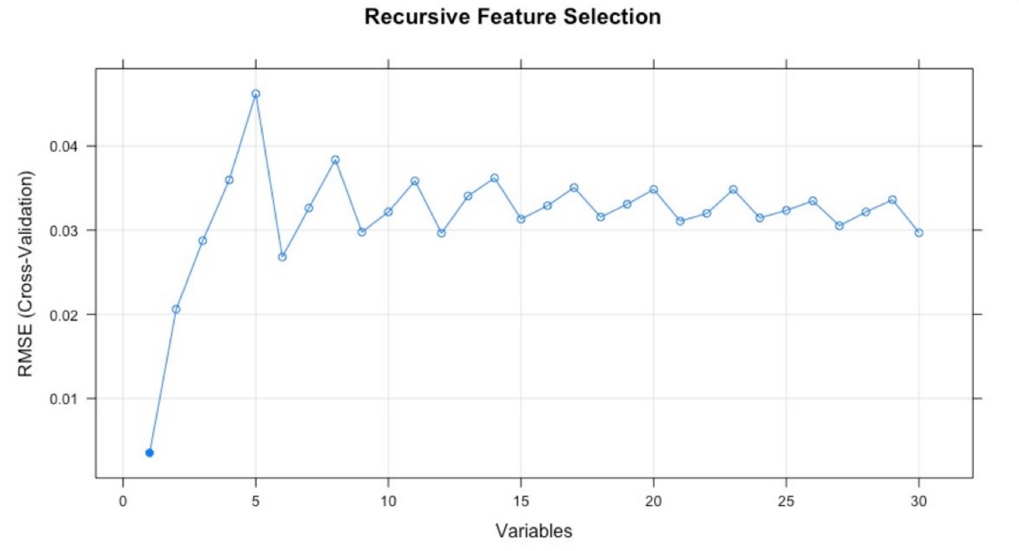


Figure 10. Recursive Feature Vector Selection Based Upon RMSE

Overall Time

The Random Forest algorithm is a computationally intensive process that requires strong computing power in order to iterate through series of trees and related decisions. Using 333 observations across 109 variables, the computation of the full model output took approximately 3 minutes and 48 seconds using a standard Macintosh desktop computer.

Conclusion/Review of Results

The purpose of the study was to develop a predictive OTL model based upon the 2014 Opportunity-to-Learn Standards for Music Instruction. The overall goal was to create a model that can accurately and efficiently map an opportunity-to-learn classification to a newly completed survey response. The following is an outline of results based upon the research questions guiding each of the three aims of the study.

Aim 1. The first aim of this study was to examine the quality of opportunity-to-learn in secondary level music performance classrooms through the development and validation of an opportunity-to-learn self-report rating scale based upon the 2014 Opportunity-to-Learn Standards for Music Instruction. The research questions that guided the first aim include:

1. *What is the psychometric quality (i.e., validity, reliability, and precision) of the rating scale used to measure opportunity-to-learn?*

   Overall, the measure of opportunity-to-learn exhibited acceptable psychometric properties, indicating that the model estimates for respondents and survey items can be meaningfully interpreted. Results of the Rasch analysis indicate that the measure has strong predictive validity as indicated by statistically significant results and high separation of reliability for the respondent facet and strong construct validity as indicated by statistically significant results and high separation of reliability for the survey item facet. The reliability of separation for respondent measures (similar to Cronbach's alpha coefficient) was 0.93, indicating that the survey items discriminated among respondents with different levels of opportunity-to-learn. The model estimates for respondent measures explained 53.69% of the variance in their responses to the survey items.

2. *How well do the survey items and domains fit the expectations of the measurement model?*

   Analysis of the residuals indicated that the model estimates were reasonable summaries of music educator responses to the opportunity-to-learn survey items, with average parameter values of Infit mean square error (MSE) and Outfit MSE of around 1.00. A total of 3 of the 112 items (2.70%) did not reasonably fit the expectations of the measurement model, suggesting that the responses to these items cannot be meaningfully interpreted. All domains reasonably fit the expectations of the measurement model, suggesting that all domain-level inferences can be meaningfully interpreted.

3. *How do the survey items and domains vary in their ability for respondents to positively agree with them?*

   The rank ordering from hardest domain for respondents to positively agree with to easiest domain for respondents to positively agree with included: (a) staffing domain, (b) materials and equipment domain, (c) facilities domain, (d) curriculum domain, and (e) scheduling domain. Rank ordering of all 112 survey items can be found in Appendix B.

Aim 2. The second aim of this study was to identify typologies of opportunity-to-learn using an unsupervised machine learning approach. The research questions that guided the second aim include:

1. *Do meaningful opportunity-to-learn typologies exist based upon systematic differential item functioning (item-by-respondent) bias indices?*

   Based upon substantive and empirical considerations, a three-cluster solution was identified as having a meaningful structure. Cluster 1 represented 34.53% (*n* = 115) of the survey respondents, Cluster 2 represented 32.43% (*n* = 108) of the survey respondents, and Cluster 3 represented 33.03% (*n* = 110) of the survey respondents.

2. *What are the predominant characteristics of the opportunity-to-learn typologies?*

   Cluster 1 was distinguished by the following characteristics:

   - Strong depth and breadth of curriculum (items 1.01, 1.02, 1.05, 1.06, 1.07, 2.05, 3.16);
   - Strong instructional considerations for students with disabilities (items 3.03, 3.04 3.11, 3.12);
   - Focus on professional development (items 3.05, 3.07, 3.09, 3.10) and music-based teacher evaluation (3.17, 3.19);
   - Administrative consideration for scheduling music classes (items 2.06, 2.07); and
   - Insufficient amount of equipment (item 4.12), materials (items 4.22), technology (items 4.12, 4.26, 4.27, 4.28, 4.30, 4.32, 4.33, 4.34) and lack in facilities (items 5.01-5.09, 5.15-5.19, 5.22-5.23).

   Cluster 2 was distinguished by the following characteristics:

   - Supported by materials and equipment (items 4.02, 4.03, 4.05, 4.06, 4.08, 4.09, 4.11, 4.13, 4.18, 4.21, 4.24, 4.25, 5.12);
   - Lack of staffing, staffing qualifications, and staff development (items 3.01, 3.02, 3.08, 3.15);
   - Scheduling and special education access concerns (items 1.09, 1.10, 2.09, 2.11, 2.12, 5.24); and
   - Teaching load concerns (items 3.31, 3.14).

   Cluster 3 was distinguished by the following characteristics:

   - Provided access to musical literature (items 4.17, 4.20);
   - Lack of funding for purchasing and maintaining instruments (items 4.01, 4.04, 4.07, 4.10, 4.15);
   - Sufficient facilities (items 5.10, 5.11, 5.13, 5.14, 5.20, 5.21);

- Lacks depth/breadth in curriculum (items 1.03, 1.04, 1.12, 1.13); and
- Staffing concerns (items 3.21, 3.22).

Aim 3. The third aim of this study was to build a Random Forest model in order to predict opportunity-to-learn classifications based upon systematic differential item functioning (respondent-by-item) bias indices. The research questions that guided the third aim include:

1. *How accurately can a Random Forest model classify music programs into each of their respective opportunity-to-learn typologies?*

   The initial model specifications consisted of three default hyperparameters (ntree, *n* = 500; mtry, *n* = 10; node size, *n* = 1) resulted in out-of-bag (OOB) error rate was 13.30%. The initial model accuracy, interpreted as the ratio of correctly predicted classifications to all possible classifications, was 87.00% (95% CI[0.81, 0.94]).

2. *What adjustments to the hyperparameters of the Random Forest model can be made to improve the model's error rate?*

   The optimized model, using the set of tuned hyperparameter specifications of ntree = 40, mtry = 8 and node size = 5, demonstrated an OOB error rate of 10.73% . The newly tuned model improved predictive performance by an accuracy rate of 2.57%.

3. *What are the most important survey items for predicting the opportunity-to-learn typologies of music programs?*

   Based on the interpretation of the plot of variable importance and recursive feature selections, five items demonstrated the most predictive efficiency toward correctly classifying each of the three opportunity-to-learn groupings: Item 1.11 (*Gifted and talented are students provided with access to special experiences according to their abilities and interests*), Item 1.03 (*Your school offers at least one alternative performing organization or emerging ensemble (e.g. jazz band) for every 450 students in the school population*), Item 3.06 (*Each school district or school provides a regular program of in-service education arranged by the district or school each year for every music educator*), Item 3.11 (*Every music educator working with special education students has received sufficient in-service training in special education*), and Item 2.06 (*Effort is made to avoid scheduling single section music classes against single section classes in other subjects*).

Future Research and Action Steps

Based on the results of this study, the following is suggested for consideration by NAfME and related stakeholders.

Pilot studies in performance evaluations: Examining the effects of opportunity-to-learn on student achievement

Further research is suggested to provide empirical evidence of the effects of OTL on student achievement. Most notably, a first step may be to examine these relationships in the context of solo and ensemble/large group performance evaluations. Currently, programs and students are measured uniformly with no considerations toward the school environment. Considering the importance of such festivals within secondary-level music programs, examination of these relationships may provide a more accurate representation of student, ensemble, and program achievement. Additionally, considerations of OTL may provide a more fair representation of teacher and program effectiveness in relation to student achievement. A strategic partnership between NAfME and state associations to gather opportunity-to-learn and performance achievement data may facilitate support within and across states and districts.

Focus groups and case studies: Examining the effects of outside variables on opportunity-to-learn

The variability of opportunity-to-learn in music classrooms may stem from a variety of sociological, financial, and other educational obstacles. Particularly in today's charged, sociological climate, it is suggested to begin an examination of the variables affecting opportunity-to-learn in music programs from the perspective of various stakeholders. Groups may consist of many types of stakeholders demonstrating support for the arts, including teachers, administrators, figures in politics, artists, and/or arts philanthropists. Many research conferences exist that highlight some of these variables; however, more focused, practitioner-based meetings between multiple stakeholders addressing specific topics inherent within the opportunity-to-learn construct may yield a better understanding of how OTL manifests across music programs.

Targeted professional development based upon opportunity-to-learn typology

The results of the study suggest that there are specific classes of opportunity-to-learn types as well as a rank ordering of OTL items. It is suggested to develop a targeted professional development plan in order to provide resources to music teachers of certain cluster types to help them overcome some of the obstacles demonstrated in the survey results. This may include providing resources for finding grants and writing grant proposals for clusters 1 and 3, providing best-practice solutions for scheduling and staffing concerns for clusters 2 and 3, or providing curricular resources for cluster 3, for example.

Focus groups: Think, pair, share from varying OTL typologies

Strengths, in addition to limitations, were identified in this study. Pairing together teachers and other stakeholders from various clusters with the purpose of examining strengths and weakness of their teaching conditions as well as best-practice solutions for overcoming OTL limitations may provide an important first step in providing systematic strategies toward overcoming limitations related to opportunity-to-learn.

Arts policy education and strategic planning

Members of the NAfME administration are incredibly adept at understanding political ramifications and resources related to arts and music education. However, the depth and breadth of this information is not often fully communicated to in-service music educators. After reviewing the clustering results of this study, the development of a strategic plan to inform music teachers, students, and other related stakeholders of their rights related to OTL conditions and systematic methods for improving OTL conditions may help facilitate teacher and district action from the bottom up.

Opportunity-to-learn reporting

The encouragement of district- and state-wide opportunity-to-learn reporting may provide better indicators of OTL patterns from a demographic perspective. Furthermore, an OTL diagnostic across districts and states may provide more transparency in the quality of support for music programs. In other words, "How well are we really doing?"

References

Abril, C. R. (2009). Responding to culture in the instrumental music program: A teacher's journey. *Music Education Research, 11*(1), 77-91.

Allsup, R. E. (2015). Music teacher quality and the problem of routine expertise. P*hilosophy of Music Education Review, 23*(1), 5-24.

Baker, R. A. (2012). The effects of high-stakes testing policy on arts education. *Arts Education Policy Review, 113*(1), 17-25.

Bautista, A., Yau, X., & Wong, J. (2017). High-quality music teacher professional development: a review of the literature. *Music Education Research, 19*(4), 455-469.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Briscoe, D. (2016). Enhancing learning for young music students: Involving and motivating parents. *Music Educators Journal, 103*(2), 41-46.

Celebi, M. E., & Aydkin, K. (2016). *Unsupervised learning algorithms*. New York, NY: Springer.

Crawford, R. (2017). Rethinking teaching and learning pedagogy for education in the twenty-first century: Blended learning in music education. *Music Education Research, 19*(2), 195-213.

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences.* New York, NY: Routledge.

Engelhard, G., Jr., & Perkins, A. F. (2011). Person response functions and the definition of units in the social sciences. *Measurement: Interdisciplinary Research & Perspective, 9*, 40-45.

Furno, M. A. (2019). *Rasch analysis of relational well-being within the National Survey of Adoptive Parents of 2007*.

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). *Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation*. University of Texas at El Paso Departmental Technical Report: UTEP-CS-18-09. Retrieved from https://scholarworks.utep.edu/cs_techrep/1209/

Google Developers. (2020). *Machine learning glossary*. Retrieved from https://developers.google.com/machine-learning/glossary/

Hasket, B. L. (2016). A survey study of U.S. collegiate and k-12 steel band directors' attitudes relating to steel band curriculum and pedagogy. *Update: Applications of Research in Music Education 34(*2), 5-12.

Heafner, T. L., & Fitchett, P. G. (2015). An opportunity to learn US history: What NAEP data suggest regarding the opportunity to learn. *The High School Journal, 98*(3), 226-249.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing interval covariate shift. *Proceedings of the 32nd International Conference on Machine Learning.* Retrieved from https://arxiv.org/abs/1502.03167

Kang, S. (2016). The history of multicultural music education and its prospects: The controversy of music universalism and its application. *Update: Applications of Research in Music Education, 34*(2), 21-28.

Kilanowski, J. F., & Lin, L. (2012). Rasch analysis of US household food security survey module in Latino migrant farmworkers. *Journal of Hunger, Environment, and Nutrition, 7*(2-3), 178-191.

Kruse, A. J. (2016). Cultural bias in testing: A review of literature and implications for music education. *Update: Applications of Research in Music Education, 35*(1), 23-31.

Latten, J. E. (1998). A scheduling-conflict resolution model. *Music Educators Journal, 84*(6), 22-26.

Lehman, P. (2014). How are we doing? In T. S. Brophy, M. L. Lai, & H. F. Chen (Eds.), *Music assessment and global diversity: Practice, measurement, and policy* (pp. 3-17). Chicago, IL: GIA Publications.

Linacre, J. M. (n.d.). *Bias interaction DIF DPF DRF estimation: Help for Facets Rasch Measurement Software*. Retrieved from www.winsteps.com/facetman/biasestimation.htm

Linacre, J. M. (1989). *Many facet Rasch measurement.* Chicago, IL: MESA Press.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328.

Linacre, J. M. (2006). Data variance explained by Rasch measures. *Rasch Measurement Transactions, 20*(1), 1045.

Linacre, J. M. (2014). *Facets*. Chicago, IL: MESA Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Mardia, K. V., Kent, J. T., & Bobby, J. M. (1979). *Multivariate analysis*. London, UK: Academic Press.

May, B. N., Willie, K., Worthen, C., & Pehrson, A. (2017). An analysis of state music education certification and licensure practices in the United States. *Journal of Music Teacher Education, 27*(1), 65-88.

May, L., & Brenner, B. (2016). The role of the arts in school reform. *Arts Education Policy Review, 117*(4), 223-229.

Menard, E. A. (2015). Music composition in the high school curriculum: A multiple case study. *Journal of Research in Music Education, 63*(1), 114-136.

Mitchell, T. (1997). *Machine learning*. New York, NY: McGraw-Hill.

Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2017). *Machine learning: Algorithms and applications.* Boca Raton, FL: CRC Press.

Moore, R. S., Brotons, M., & Jacobi-Kama, K. (2002). Observational analysis of instructional time in general music: Comparison of U.S.A. And Spanish teachers in grades k-5. *Bulletin of the Council for Research in Music Education, 153/154*, 48-54.

Morrison, S. J. (2001). The school ensemble: A culture of our own. *Music Educators Journal, 88*(2), 24-28.

Music Educators National Conference (1994). *Opportunity-to-learn standards for music instruction*. Reston, VA: MENC.

Music Educators National Conference (1999). *Opportunity-to-learn standards for music technology*. Reston, VA: MENC.

National Association for Music Education (NAfME) (2014). *Opportunity-to-learn standards*. Retrieved from https://nafme.org/wp-content/files/2014/11/Opportunity-to-Learn Standards_May2015.pdf

National Council on Education Standards and Testing (1992). *Raising standards for American education: A report to congress, the Secretary of Education, the National Education Goals Panel, and the American people*. Washington, DC: U.S. Government Printing Office.

Olsen, R. V., Garratt, A. M., Iversen, H. H., & Bjertnaes, O. A. (2010). Rasch analysis of the Psychiatric Out-Patient Experiences Questionnaire (POPEQ). B*MC Health Services Research, 10*.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* MESA Press.

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics, 4*(3), 207–230.

Rogers, G. L. (2016). The music of the spheres: Cross-curricular perspectives on music and science. *Music Educators Journal, 103*(1), 41-48.

Romesburg, H. C. (1984). *Cluster analysis for researchers.* Belmont, CA: Lifetime Learning.

Rousseau, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics, 20*(1), 53-65.

Russell, J. A., & Austin, J. R. (2010). Assessment practices of secondary music teachers. *Journal of Research in Music Education, 58(*1), 37-54.

Salvador, K., & Kelly-McHale, J. (2017). Music teacher educator perspectives on social justice. *Journal of Research in Music Education, 65*(1), 6-24.

Schmidt, M., & Smith, M. (2017). Creating culturally responsive ensemble instruction: A beginning music educator's story. *Bulletin of the Council for Research in Music Education, 111/112*, 61-79.

Schmidt, W.H., & Maier, A. (2009). Opportunity to learn. In G. Sykes, B. L. Schneider, & D. N. Plank (Eds.), *Handbook on education policy research* (pp. 541-549). New York: Routledge.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms.* New York, NY: Cambridge University Press.

Shaw, R. D. (2016). Arts teacher evaluation: How did we get here? *Arts Education Policy Review, 117*(1), 1-12.

Shuler, S. C., Brophy, T. S., Sabol, R., McGreevy-Nichols, S., & Schuttler, M. J. (2016). Arts assessment in the age of accountability: Challenges and opportunities in implementation, design, and measurement. In H. Braun (Ed.), *Meeting the challenges to measurement in an era of accountability* (pp. 183–216). New York, NY: Routledge.

Sotiropoulou-Zormpala, M. (2016). Seeking a higher level of arts integration across the curriculum. *Arts Education Policy Review, 117*(1), 43-54.

Stevens, F. I., & Grymes, J. (1993). *Opportunity to learn: Issues of equity for poor and minority students*. Washington, DC: National Center for Education Statistics.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika, 18*(4), 267–276.

U.S. Department of Education. (2010). *The condition of education, 2010* (NCES Report No. 2010-028, Indicator 21). Washington, DC: National Center for Education Statistics.

U.S. Government (1991). *Education Council Act of 1991* § 102, 62 Stat. 305 (1991). Retrieved from https://www.congress.gov/

U.S. Government. (1993). H . R . 1804 *Goals 2000 : Educate America Act*. Congressional Record, 1–9. Retrieved from https://www.congress.gov/

Vagias, W. M. (2006). *Likert-type scale response anchors*. Clemson, SC: Clemson University.

Valli, L., Cooper, D., & Frankes, L. (1997). Professional development school and equity: A critical analysis of rhetoric and research. In M. W. Apple (ed.) *Review of Research in Education, Volume 22*. Washington DC: American Educational Research Association.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.

Wesolowski, B. C. (2019). Item Response Theory in music testing. In T. Brophy (Ed.), T*he Oxford handbook of assessment policy and practice in music education* (pp. 479-503). New York: Oxford University Press.

Wesolowski, B. C., Wind, S. A., & Engelhard, Jr., G. (2015). Rater fairness in music performance assessment: Evaluating model-data fit and differential rater functioning. *Musicae Scientiae, 19*(2), 147-170.

Wright, B. D. (2000). Rasch analysis for surveys. *Popular Measurement, 3*(1), 61.

Wright, B. D., & Linacre, J. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis.* Chicago: IL: MESA Press.

APPENDIX A

112-Item Opportunity-to-Learn Rating Scale

Item	Domain	Criteria	1	2	3	4	5
1.01	Curriculum	The school curriculum adequately covers all three artistic processes (Performing, Responding, and Creating) aligned with the 2014 National Association for Music Education (NAfME) Standards.	Never	Rarely	Occasionally	Often	Always
1.02	Curriculum	The school curriculum allows for the transfer of all three artistic processes (Performing, Responding, and Creating) by adequately aligning the Connecting Standards found witching the 2014 NAfME Standards.	Never	Rarely	Occasionally	Often	Always
1.03	Curriculum	Your school offers at least one alternative performing organization or emerging ensemble (e.g. jazz band) for every 450 students in the school population.	No	Yes			
1.04	Curriculum	Ensembles are differentiated by experience and age level when justified by enrollment.	Never	Rarely	Occasionally	Often	Always
1.05	Curriculum	Music students are offered small group instruction with a focus on improvisation.	No	Yes			
1.06	Curriculum	Instruction is available for those students interested in addressing new experiences in Ensembles and Harmonizing Instruments at the Novice or Intermediate levels.	No	Yes			
1.07	Curriculum	String program instruction begins no later than grade 4.	No	Yes			
1.08	Curriculum	Band program instruction begins no later than grade 5.	No	Yes			
1.09	Curriculum	Students with special needs are given the same opportunities to elect musical instruction as other students.	Never	Rarely	Occasionally	Often	Always
1.10	Curriculum	Special education classes in music are no larger than other special education classes.	No	Yes			
1.11	Curriculum	Gifted and talented are students provided with access to special experiences according to their abilities and interests.	Never	Rarely	Occasionally	Often	Always
1.12	Curriculum	The program provides all students the opportunity to achieve at levels consistent with their individual abilities.	Strongly Disagree	Disagree	Agree	Strongly Agree	
1.13	Curriculum	The program provides all students the opportunity to achieve at levels consistent with the 2014 NAfME Standards listed for the appropriate grade levels.	Strongly Disagree	Disagree	Agree	Strongly Agree	

2.01	Scheduling	Every performing group presents a series of performances for parents, peers, or the community.	No	Yes			
2.02	Scheduling	At least one performing group of each type such as band, chorus, orchestra, or guitar performs once yearly at a premiere venue. (The venue may be local or involve travel out of the school district.)	No	Yes			
2.03	Scheduling	The amount of performances reduces the amount of time available to achieve the instructional objectives of the ensemble.	Strongly Disagree	Disagree	Agree	Strongly Agree	
2.04	Scheduling	The performance schedule suggests a focus on entertainment rather than education.	Never	Rarely	Occasionally	Often	Always
2.05	Scheduling	Instruction in ensembles is commensurate with other core subject areas.	Never	Rarely	Occasionally	Often	Always
2.06	Scheduling	Effort is made to avoid scheduling single section music classes against single section classes in other subjects.	Strongly Disagree	Disagree	Agree	Strongly Agree	
2.07	Scheduling	Ensembles and large music classes are offered at times designed to allow participation by the maximum number of students.	Strongly Disagree	Disagree	Agree	Strongly Agree	
2.08	Scheduling	Scheduling is arranged so that all members of each ensemble can meet as a unit each school day.	Never	Rarely	Occasionally	Often	Always
2.09	Scheduling	Students in performance ensembles are scheduled by experience and/or student proficiency level.	No	Yes			
2.10	Scheduling	Performing ensemble classes do not interfere with student participation in general music classes.	Never	Rarely	Occasionally	Often	Always
2.11	Scheduling	Pullouts for school assemblies, test preparation, or other non-music education activities are arranged to minimally impact music learning.	Strongly Disagree	Disagree	Agree	Strongly Agree	
2.12	Scheduling	Just as other core academic subject areas meet during the course of the curricular school day, after-school rehearsals serve to supplement the learning that takes place within the school day.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.01	Staffing	Instruction is provided by Highly Qualified/Certified music teachers who have received formal training (including in-service training) in the ensemble taught.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.02	Staffing	Certified non-arts educators are drawn on to expand students' opportunities for arts learning by providing curricular connections among the arts and other subjects.	Never	Rarely	Occasionally	Often	Always

3.03	Staffing	Teacher aides are provided for special education classes in music if they are provided for other special education classes.	Never	Rarely	Occasionally	Often	Always
3.04	Staffing	If a student with a disability has an aide to assist in other classes, the aide also assists the student in music classes.	Never	Rarely	Occasionally	Often	Always
3.05	Staffing	Teachers have regular access to professional development materials and experiences in their performance area, including online NAfME resources.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.06	Staffing	Each school district or school provides a regular program of in-service education arranged by the district or school each year for every music educator.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.07	Staffing	Each school district or school provides at least two paid days for professional development activities arranged by the district or school each year for every music educator.	No	Yes			
3.08	Staffing	Every music educator is permitted at least one paid day of leave each year for professional development activities proposed by the teacher and approved by the school.	No	Yes			
3.09	Staffing	Music staff members are encouraged and supported to participate in state and national professional development events.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.10	Staffing	Music staff are supported and encouraged to assume leadership roles in state and national music organizations.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.11	Staffing	Every music educator working with special education students has received sufficient in-service training in special education.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.12	Staffing	For purposes of consultation, every music educator working with special education students, has convenient access to trained professionals in special education and/or music therapy.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.13	Staffing	Class loads for music teachers are not significantly higher than other academic areas.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.14	Staffing	Student to teacher ratios should be established to ensure additional music teachers are hired to ensure equitable music instruction for all students.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.15	Staffing	Time is provided for collaborative work groups/professional learning communities.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.16	Staffing	Teacher evaluation is carried out in a way consistent with that of teachers in other subjects, except that the provisions of the NAfME Position Statement on Teacher Evaluation are met (notably, the use of student outcome measures is limited to student achievement in music).	Strongly Disagree	Disagree	Agree	Strongly Agree	

3.17	Staffing	Teacher evaluation includes a balanced, comprehensive assessment of the teacher's contributions to student learning through multiple measures. These measures can and should collect information such as: * Indicators of teacher practice, such as planning and preparation. * Indicators of the teacher's role in maintaining a productive classroom environment. * Indicators that instruction is designed to reach specified goals related to the Artistic Processes of Creating, performing, and Responding, as well as to the "connecting" embedded in those processes. * Indicators of teacher contribution to the school or district, as well as to the profession of teaching at large.	Strongly Disagree	Disagree	Agree	Strongly Agree	
3.18	Staffing	If student performance data are considered in teacher evaluation, data must involve music outcomes.	No	Yes			
3.19	Staffing	Teacher evaluation is conducted by individuals fully qualified in both evaluation and music instruction.	No	Yes			
3.20	Staffing	One music educator in every district or school is designated as coordinator or administrator to provide leadership for the music program(s).	No	Yes			
3.21	Staffing	Coordinator (see question 3.20) is employed on a full-time basis for administration when the staff includes twenty-five or more music educators.	No	Yes			
3.22	Staffing	The amount of administrative time is adjusted proportionately to the size of the staff.	No	Yes			
3.23	Staffing	Additional administrative staff is employed at a proportional rate when the staff is larger.	No	Yes			
4.01	Materials & Equipment	Instruments are provided where students have difficulty in purchasing instruments due to financial hardship.	Never	Rarely	Occasionally	Often	Always
4.02	Materials & Equipment	The instruments listed below are provided in sufficient quantity. *Middle School Band: C piccolos, bass clarinets, tenor saxophones, baritone saxophones, oboes, bassoons, double French horns, baritone horns, tubas, concert snare drums, pedal timpani, concert bass drums, crash cymbals, suspended cymbals, tambourines, triangles, xylophones and marimbas, orchestral bells, assorted percussion equipment. *Middle School Jazz Ensemble: in addition to listings for Middle School Band, baritone sax, electric bass with amplifier, trap set. *High School Strings: same as Middle School Strings. *High School Jazz Ensemble: in addition to listings for Middle School Jazz Ensemble, bass trombone. * High School Band: in addition to listings for Middle School Band, E-flat clarinets, A clarinets, alto clarinets, contrabass clarinets, bass trombones. *Emerging Ensembles: guitars, drums, pans, as appropriate for the ensemble.	None	Very Little	Some	Quite a bit	Very much

4.03	Materials & Equipment	Every room in which music is taught has convenient access to a high-quality acoustic or electronic piano.	No	Yes			
4.04	Materials & Equipment	Instruments are maintained in good repair.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.05	Materials & Equipment	An annual budget is provided for repair and maintenance of instruments equal to at least 5% of the current replacement value of the total inventory of instruments and equipment.	No	Yes			
4.06	Materials & Equipment	The school program has a written depreciation and replacement plan for all instruments and equipment, specifically describing under what conditions instruments should be retired and replaced.	No	Yes			
4.07	Materials & Equipment	All instruments supplied by the school are of a quality generally understood to be that of undamaged "student line" instruments, and thus are appropriate for student learning and performance.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.08	Materials & Equipment	All instruments provided by the school exceed the quality generally understood to be that of undamaged "student line" instruments, and thus are appropriate for more advanced student learning and performance.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.09	Materials & Equipment	There are funds available to purchase several higher quality instruments (college level) for advanced students.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.10	Materials & Equipment	Instruments are provided to develop emerging ensembles and classes, including non-traditional or non-western instruments (These could include many different instruments such as steel drums, iPads, West-African drums, and Chinese erhus).	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.11	Materials & Equipment	Accessories (conductor's stands, tuning stands, music folders, chairs designed for music classes, drum stands, moveable percussion cabinets, tuba chairs, bass stools) are provided in sufficient quantity.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.12	Materials & Equipment	Every room in which music is taught has sufficient sturdy music stands.	No	Yes			
4.13	Materials & Equipment	Adaptive devices (such as adaptive picks, beaters) are available for use by students with special needs.	No	Yes			
4.14	Materials & Equipment	If a music task cannot be performed by students with special needs exactly as it would be by other students, adaptation is provided so that students with special needs can participate insofar as possible.	Strongly Disagree	Disagree	Agree	Strongly Agree	

4.15	Materials & Equipment	There is a budget available for specialized music accessories, as needed.	No	Yes			
4.16	Materials & Equipment	There is a folder of original music for each stand of no more than two instrumentalists.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.17	Materials & Equipment	Original copies of music are provided for each student for instruments where sharing stands is not feasible or traditional.	No	Yes			
4.18	Materials & Equipment	The music library contains music appropriate for various levels from which students may choose.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.19	Materials & Equipment	The music library contains no materials produced in violation of copyright laws.	No	Yes			
4.20	Materials & Equipment	The music library contains at least 75 titles/musical works for each type of performing group.	No	Yes			
4.21	Materials & Equipment	At least 5 titles/musical works for each performing group are added to the music library each year.	No	Yes			
4.22	Materials & Equipment	At least 15 titles/musical works are added to the music library each year.	No	Yes			
4.23	Materials & Equipment	Main school library/media center/resource center contains a variety of music-related books.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.24	Materials & Equipment	Annual budget is provided for supplies including: recordings/downloads, computer media, and other special supplies, materials and equipment needed for teaching the music curriculum.	No	Yes			
4.25	Materials & Equipment	School has technology available for music instruction (computers and appropriate software, including notation, sequencing, and audio editing software; printers, audio and video input and output devices, electronic keyboards).	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.26	Materials & Equipment	Equipment is provided that keeps pace with changing technologies and needs.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.27	Materials & Equipment	Every room in which music is taught is equipped with a high-quality sound reproduction system capable of using current recording technology.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.28	Materials & Equipment	Every room in which music is taught is equipped with a high-quality video reproduction system capable of using current recording technology.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.29	Materials & Equipment	At least some of the audio equipment can be operated by students.	Strongly Disagree	Disagree	Agree	Strongly Agree	

4.30	Materials & Equipment	Each ensemble has available at least one electronic version of key ensemble instruments so that students can gain experience with these instruments (e.g., electric violin, MIDI wind controller, electric guitar).	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.31	Materials & Equipment	Teachers have easy access to email, online storage, a school-sanctioned web portal and other online services for professional and curricular development, research, and other communications needs.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.32	Materials & Equipment	Teachers have access to quality projectors and/or interactive boards.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.33	Materials & Equipment	Every teacher has convenient access to sound recordings representing a wide variety of music styles and cultures.	Strongly Disagree	Disagree	Agree	Strongly Agree	
4.34	Materials & Equipment	Technology is available to support student assessment strategies adopted by the school or district.	Strongly Disagree	Disagree	Agree	Strongly Agree	
5.01	Facilities	Spaces used for music instruction are adequate in size to accommodate the largest group taught.	No	Yes			
5.02	Facilities	Instrumental rehearsal rooms contain at least 1,800 sq. ft. of floor space.	No	Yes			
5.03	Facilities	Instrumental rehearsal rooms contain at least 2,500 sq. ft. of floor space.	No	Yes			
5.04	Facilities	Spaces used for music instruction have appropriate acoustical properties. Each room is acoustically isolated by an acoustical barrier or wall with a Sound Transmission Classification (STC) of 50 or more.	No	Yes			
5.05	Facilities	Instrumental rehearsal room contains a ceiling at least 16 ft. high.	No	Yes			
5.06	Facilities	Instrumental rehearsal room contains a ceiling at least 20 ft. high.	No	Yes			
5.07	Facilities	Instrumental rehearsal room contains a double entry door.	No	Yes			
5.08	Facilities	Instrumental rehearsal room ventilation provides air exchange rate double that of an ordinary classroom.	No	Yes			
5.09	Facilities	Lighting and ventilation systems do not exceed Noise Criterion levels of 20 for auditoria or other rooms designated for performances, and 30 for classrooms, rehearsal rooms, and practice rooms or studios.	No	Yes			
5.10	Facilities	School contains at least one practice room of at least 55 sq. ft. for each 40 students enrolled in performing groups.	No	Yes			
5.11	Facilities	School contains at least one practice room of at least 55 sq. ft. for each 20 students enrolled in performing groups.	No	Yes			

5.12	Facilities	Individual areas, with access to recording equipment, are provided for the purpose of student assessment.	No	Yes			
5.13	Facilities	Office or studio space is provided to each music educator adjacent to the instructional area in which the educator teaches.	No	Yes			
5.14	Facilities	Office or studio space has convenient access to telephone and internet-connected computer.	No	Yes			
5.15	Facilities	Space is available for the repair and maintenance of instruments.	No	Yes			
5.16	Facilities	A space available for repairs has access to running water.	No	Yes			
5.17	Facilities	Sufficient secured storage space is available to store instruments, equipment, and instructional materials.	Strongly Disagree	Disagree	Agree	Strongly Agree	
5.18	Facilities	Cabinets and shelving are provided, as well as lockers for the storage of instruments in daily use.	No	Yes			
5.19	Facilities	The storage space is immediately adjacent to the rehearsal facilities.	No	Yes			
5.20	Facilities	Separate spaces are provided for music instruction and music performance.	No	Yes			
5.21	Facilities	Performance venues are adequate to accommodate the largest group taught.	No	Yes			
5.22	Facilities	Performance venues have appropriate properties of acoustics, lighting, secure storage, and sound.	No	Yes			
5.23	Facilities	Students have access to high quality performance venues at least once a year to enable them to present academic accomplishments to the public.	No	Yes			
5.24	Facilities	At least one performance venue is available that provides seating for the entire school population.	No	Yes			

APPENDIX B

Calibration of Opportunity-to-Learn Survey Items

Calibration of Opportunity-to-Learn Survey Items

Item	Observed Average	Measure	Standard Error	Infit MSE	Standardized Outfit	Outfit MSE	Standardized Outfit
4.06	1.12	2.14	0.17	0.96	-0.26	0.88	-0.74
5.11	1.13	2.02	0.16	0.93	-0.56	0.87	-0.94
5.12	1.19	1.53	0.14	0.88	-1.57	0.78	-2.26
3.11	1.68	1.50	0.08	1.07	0.97	1.05	0.73
5.03	1.21	1.39	0.13	0.95	-0.73	0.93	-0.71
3.19	1.53	1.38	0.08	1.04	0.46	1.11	1.17
1.07	1.24	1.28	0.14	1.05	0.78	1.19	1.99
4.28	1.50	1.28	0.08	0.97	-0.25	1.03	0.29
3.22	1.24	1.23	0.13	0.94	-0.95	0.94	-0.73
4.27	1.61	1.23	0.07	0.94	-0.70	1.00	-0.01
2.04	1.77	1.16	0.08	1.53	5.89	1.71	7.65
4.09	1.25	1.16	0.12	0.87	-2.19	0.80	-2.70
1.05	1.26	1.12	0.12	0.96	-0.63	0.94	-0.84
4.13	1.26	1.11	0.12	0.93	-1.18	0.86	-1.89
5.10	1.28	1.00	0.12	0.97	-0.51	0.95	-0.76
3.14	2.01	0.89	0.07	1.06	0.89	1.05	0.82
4.10	1.93	0.88	0.07	0.90	-1.48	0.89	-1.64
3.21	1.31	0.87	0.12	0.96	-0.76	0.96	-0.72
3.06	1.87	0.82	0.06	1.03	0.49	1.07	0.97
3.18	1.77	0.81	0.06	1.39	4.84	1.47	4.45
3.02	1.98	0.78	0.08	1.03	0.39	1.04	0.55
4.23	2.40	0.75	0.07	0.98	-0.20	0.97	-0.35
5.17	1.93	0.71	0.06	1.04	0.55	1.05	0.63
5.04	1.34	0.70	0.11	0.93	-1.81	0.90	-1.89
5.06	1.34	0.70	0.11	0.97	-0.81	0.97	-0.66
4.22	1.34	0.69	0.11	0.96	-0.97	0.94	-1.07
3.13	2.13	0.68	0.06	1.20	3.13	1.22	3.36
5.16	1.36	0.64	0.11	0.99	-0.34	0.98	-0.41
4.15	1.36	0.63	0.11	0.89	-2.95	0.85	-3.14
4.25	2.52	0.57	0.06	0.84	-2.50	0.83	-2.61
5.08	1.37	0.57	0.11	0.96	-1.16	0.94	-1.26
4.26	2.24	0.55	0.07	0.81	-3.25	0.81	-3.15
4.16	2.10	0.52	0.07	1.08	1.09	1.09	1.29
2.03	2.17	0.50	0.08	1.44	5.00	1.45	5.06
5.15	1.39	0.50	0.11	0.97	-0.94	0.95	-1.12
4.05	1.40	0.43	0.11	0.90	-3.03	0.89	-2.81
3.20	1.41	0.40	0.11	1.01	0.25	1.02	0.62
1.10	1.41	0.39	0.13	1.07	1.86	1.07	1.50
3.03	2.11	0.37	0.06	1.05	0.80	1.05	0.76

3.17	2.27	0.37	0.06	1.23	3.42	1.24	3.51
3.12	2.30	0.36	0.07	0.92	-1.27	0.92	-1.30
2.11	2.34	0.30	0.06	1.09	1.36	1.12	1.82
5.22	1.43	0.28	0.11	0.96	-1.38	0.96	-1.17
4.08	2.33	0.24	0.07	0.88	-1.76	0.89	-1.70
3.04	2.28	0.17	0.06	1.07	1.09	1.08	1.19
1.08	1.47	0.14	0.11	1.12	4.24	1.15	4.26
2.10	2.35	0.13	0.05	1.46	6.97	1.62	7.79
3.15	2.55	0.08	0.07	0.92	-1.27	0.91	-1.27
5.09	1.48	0.07	0.11	0.94	-2.10	0.93	-2.24
1.06	1.49	0.03	0.11	0.98	-0.64	0.98	-0.65
1.04	2.51	0.01	0.06	1.12	1.80	1.12	1.87
4.24	1.50	-0.02	0.11	0.87	-5.13	0.85	-4.72
2.06	2.61	-0.03	0.06	1.00	0.05	1.00	0.05
1.11	2.47	-0.04	0.07	1.00	0.00	1.00	-0.01
5.24	1.51	-0.05	0.11	1.12	4.35	1.12	3.59
4.11	2.66	-0.09	0.07	0.78	-3.54	0.77	-3.57
2.08	2.69	-0.18	0.05	1.21	3.33	1.27	3.60
5.02	1.54	-0.18	0.11	0.95	-1.95	0.93	-2.17
4.07	2.76	-0.20	0.07	0.92	-1.09	0.92	-0.98
3.16	2.77	-0.23	0.07	1.07	0.90	1.11	1.37
3.07	1.56	-0.25	0.11	0.96	-1.36	0.95	-1.34
5.01	1.56	-0.25	0.11	0.95	-1.87	0.93	-2.05
2.05	2.67	-0.27	0.06	1.17	2.61	1.20	2.97
2.09	1.56	-0.28	0.11	1.02	0.72	1.03	0.79
4.14	2.79	-0.38	0.07	0.93	-1.03	0.93	-0.98
4.03	1.60	-0.45	0.11	1.00	-0.01	0.99	-0.35
2.07	2.87	-0.46	0.07	0.95	-0.71	0.94	-0.77
4.04	2.85	-0.46	0.08	0.89	-1.38	0.91	-1.11
4.32	2.83	-0.46	0.07	0.92	-1.08	0.94	-0.88
4.31	2.88	-0.48	0.07	0.91	-1.28	0.91	-1.26
1.03	1.62	-0.53	0.11	0.94	-1.61	0.93	-1.55
3.10	2.81	-0.53	0.07	0.91	-1.36	0.90	-1.42
5.07	1.62	-0.53	0.11	1.02	0.50	1.01	0.14
4.30	3.01	-0.60	0.07	1.06	0.72	1.08	0.95
4.17	1.64	-0.62	0.11	1.03	0.70	1.03	0.73
1.09	2.79	-0.63	0.07	1.17	2.50	1.19	2.77
5.05	1.65	-0.65	0.11	0.98	-0.60	0.97	-0.67
1.01	2.84	-0.67	0.08	1.04	0.54	1.04	0.55
4.02	3.42	-0.69	0.06	0.96	-0.64	0.97	-0.42
1.02	2.82	-0.72	0.08	1.03	0.53	1.03	0.51
1.13	2.94	-0.74	0.08	0.90	-1.18	0.89	-1.38
5.21	1.67	-0.75	0.11	0.99	-0.27	0.99	-0.21
4.21	1.67	-0.76	0.11	0.92	-1.83	0.90	-1.94
4.19	1.68	-0.80	0.12	1.05	1.17	1.06	1.07
2.12	3.07	-0.82	0.08	1.11	1.38	1.13	1.57
3.09	2.97	-0.82	0.07	0.92	-1.22	0.91	-1.36
4.20	1.68	-0.82	0.12	1.00	0.02	1.00	0.01

4.01	2.92	-0.84	0.06	1.03	0.52	1.04	0.69
3.05	3.10	-0.92	0.08	0.95	-0.63	0.92	-0.97
5.18	1.70	-0.92	0.12	0.94	-1.18	0.93	-1.18
5.13	1.71	-0.95	0.12	0.94	-1.13	0.92	-1.23
3.08	1.71	-0.98	0.12	1.00	-0.03	1.03	0.47
4.18	3.15	-0.99	0.08	0.91	-1.14	0.92	-1.05
2.02	1.72	-1.00	0.12	1.05	0.91	1.10	1.47
1.12	3.11	-1.05	0.09	0.87	-1.61	0.85	-1.81
5.23	1.73	-1.05	0.12	0.92	-1.40	0.88	-1.72
4.12	1.75	-1.19	0.12	0.90	-1.71	0.82	-2.45
5.20	1.77	-1.28	0.13	0.96	-0.52	0.90	-1.20
4.29	3.37	-1.50	0.09	0.94	-0.73	0.91	-1.07
3.01	3.62	-1.64	0.09	1.09	0.84	1.11	0.93
5.14	1.83	-1.73	0.14	0.95	-0.50	0.89	-0.91
5.19	1.85	-1.88	0.15	0.97	-0.31	0.96	-0.31
2.01	1.97	-3.74	0.32	0.99	0.06	0.97	0.03

Note. Survey items are presented in measure order from the hardest survey item for respondents to agree with to the easiest survey item for respondents to agree with.