Examining the Psychometric Quality of a Modified Perceived Authenticity in Writing Scale with Rasch Measurement Theory

Nadia Behizadeh and George Engelhard Jr.

High-stakes, large-scale testing has proliferated in the United States, and a plethora of studies indicate that instructional practices have suffered (e.g., Darling-Hammond 2010). In particular, researchers theorize that although US students are writing more (Applebee and Langer 2011), classroom writing experiences are not highly authentic to students, especially for urban, students of color who are economically disadvantaged (Ball and Ellis 2008). Authenticity is a key motivational variable in school settings, and if US students are not experiencing authentic writing instruction, then reasons for this lack, such as the potential negative effects of current large-scale writing assessments, need to be addressed. Alternately, if some students are experiencing highly authentic writing instruction, a closer examination of factors that contribute to the enactment of authentic writing instruction needs to occur. However, because authenticity is a student's perception of the meaningfulness of instruction, student perspectives are needed to explore the authenticity of writing instruction in the United States.

There are not many tools available for measuring authenticity. One past tool is the Perceived Authenticity in Writing (PAW) Scale designed to measure perceived authenticity in writing instruction for adolescents for a *specific task* (Behizadeh and Engelhard 2014). However, there is a need for a similar scale to the PAW Scale, but one that could be used for examining students' general impression of their writing instruction as a whole. This would allow researchers, educators, and policymakers

N. Behizadeh (✉)
Department of Middle and Secondary Education, College of Education
and Human Development, Georgia State University, P.O. Box 3978, Atlanta,
GA 30302-3978, USA
e-mail: nbehizadeh@gsu.edu

G. Engelhard Jr.
Educational Psychology, The University of Georgia, Aderhold Hall, Athens,
GA 30602, USA
e-mail: gengelh@uga.edu

71

to administer the instrument across a multitude of contexts, and then (1) compare perceptions of authenticity; (2) identify schools or districts with high or low authenticity for deeper qualitative examination; and (3) analyze correlations among authenticity and other variables, such as socioeconomic status. In particular, this last purpose for a general authentic writing scale would allow researchers to identify differential access to authentic writing instruction and to explore issues related to social justice in writing assessment.

To this end, a modified instrument was created: the Modified Perceived Authenticity in Writing (MPAW) Scale. The MPAW Scale asks students to evaluate their overall impression of the authenticity of the current writing instruction that they are receiving. This chapter examines the psychometric properties of the MPAW Scale for use in a larger study of perceived authenticity among urban students of color in the US. The following research questions guided this study:

1. Does the internal structure of the MPAW Scale represent gradations of item difficulty?
2. Do the MPAW items exhibit acceptable model-data fit that supports the validity of inferences regarding student perceptions of the authenticity of their writing instruction?
3. Do the MPAW items exhibit measurement invariance when explored with explanatory variables such as grade level, gender, and student attitude toward writing?

Literature Review

What Is Authenticity?

Although numerous scholars call for increasing the authenticity of literacy education and writing education in particular, the meaning of "authentic" is somewhat unclear. In past research, educational authenticity has traditionally been defined as the connection of a school task to the real world (Newmann et al. 1996; Purcell-Gates et al. 2012; Seunarinesingh 2010). However, drawing on the idea of authenticity as subjective (Ashton 2010; Splitter 2009), past research (Behizadeh 2014, 2015) has presented a definition of authenticity in writing as a student's *perception* that a writing task connects to their life. This perception of authenticity includes culture, personal interests, and community or global issues that matter to the student. Importantly, this definition establishes that the authority for determining authenticity resides in the student, not with teachers or policymakers. Educators and researchers may hypothesize that particular tasks are highly authentic for students, but without confirmation from students that these tasks are indeed meaningful and relevant to their lives, a strong claim for authenticity cannot be made.

Why Does Authenticity in Writing Matter?

A large body of research supports authenticity in education as a key component for increasing student engagement and achievement, particularly in teaching writing (Fisher 2007; Freire 1970/2000; Morrell 2008; Purcell-Gates et al. 2007; Sisserson et al. 2002; Winn and Johnson 2011). In Hillocks' (2011) review of a century of literacy research, he stated, "We know from a very wide variety of studies in English and out of it, that students who are authentically engaged with the tasks of their learning are likely to learn much more than those who are not" (p. 189). Across literacy research connected to authenticity, there is the belief that greater authenticity increases student engagement and achievement. In essence, perceived authenticity by students can serve an important motivational role in educational settings. Additionally, standards for English language arts stress the importance of "authentic, open-ended learning experiences" (National Council of Teachers of English and International Reading Association 2012, p. 6) for student learning, as do documents outlining twenty-first century skills (Partnership for twenty-first Century Learning n.d). Additionally, the Common Core State Standards for English language arts (Common Core State Standards Initiative 2015) emphasize writing for real audiences and publishing and distributing student writing. Finally, connecting instructional content to the real world is a consideration of standards used to evaluate teachers (Council of Chief State School Officers 2011).

What Factors May Be Impeding Authenticity in Writing Instruction?

In addition to wide support for the importance of authenticity in writing instruction, literacy researchers have documented the misalignment between writing assessments that focus on conventions and mechanics and a definition of writing as an iterative, social, and creative contextualized process (Au and Gourd 2013; Dyson and Freedman 2003). According to leading assessment scholars, "The overreliance on psychometric approaches to assessment risks reducing diversity in teaching, learning, and assessment practices; dismissing alternative disciplinary experiences; and marginalizing local knowledge and expertise" (Haertel et al. 2008, p. 77). Thus, there is a conflict between high-stakes writing assessments that encourage rote writing instruction and research and standards supporting meaningful, authentic writing instruction.

One way to position an argument for examining the authenticity of writing instruction in relation to writing assessment is through a validity lens (Messick 1995; Kane 2013). According to Messick and Kane, validity is the use of a test for a particular purpose (in this case, to evaluate writing achievement), and validity is

evaluated by developing an argument that includes multiple sources of evidence. Messick (1995) offered a unified theory of validity, stating, "Validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment" (p. 741). Condensing Messick's (1995) six connected aspects of validity to two, major sources of validity evidence are (1) the match between the theorized construct of the assessment (definition, processes, structural elements) and the representation of this construct in the assessment; and (2) the consequences of assessment. The degree to which assessment practices impact instruction, including positive or negative effects on perceived authenticity, are part of consequential validity. However, construct and consequential validity are related. Slomp et al. (2014) articulated that issues with construct validity, especially misrepresenting or under representing the construct of writing, are closely connected to issues with consequential validity. Applying this to authenticity, if assessments are based on the idea of writing as a decontextualized set of skills, then this construction of writing can lead to teaching practices that focus on building skills without attending to the epistemic and identity-related aspects of writing.

Because of the importance of consequential validity in evaluating a writing assessment, an instrument for measuring the perceived authenticity of writing instruction by students is a tool that can be a useful for collecting validity evidence. If particular assessments are impeding authentic writing instruction, these assessments may need to be revised based on data collected from instruments such as the MPAW Scale.

Is There Evidence that Large-Scale Writing Assessments Are Reducing Authenticity?

A wide range of research indicates that high-stakes, large-scale writing assessment is impeding authentic writing instruction. In a qualitative study, Luna and Turner (2001) interviewed teachers administering high-stakes writing tests in Massachusetts, and the authors reported that teachers felt they were teaching to the test instead of providing rich writing instruction. A focus on ensuring students learned the formula for the five-paragraph essay rather than effective communication was critiqued as a negative outcome of high-stakes writing tests. In another study conducted in North Carolina, researchers found that high-stakes writing assessment resulted in "form over content and product over process" (Watanabe 2007, cited in Au and Gourd 2013, p. 17). Similarly, in their review of writing research, Dyson and Freedman (2003) argued that quality of writing depends on students' investment in a topic and their need to communicate information,

constituting a validity problem for standardized writing tests that are not compelling to students. Furthermore, negative washback is more pronounced for culturally and linguistically diverse students (Ball and Ellis 2008; Madaus 1994). In Ball and Ellis' (2008) review of decades of writing research, they concluded "that students of color are disproportionately relegated to classrooms using drill exercises rather than interactive, meaningful approaches that require extended writing, reflection, and critical thinking" (p. 507). However, although researchers can hypothesize that instruction is not authentic to students, an instrument that collects students' judgments of authenticity could help examine the degree to which certain assessments are affecting authenticity.

Methods

Again, our research questions are: (1) Does the internal structure of the MPAW Scale represent gradations of item difficulty? (2) Do the MPAW items exhibit acceptable model-data fit that supports the validity of inferences regarding student perceptions of the authenticity of their writing instruction? and (3) Do the MPAW items exhibit measurement invariance when explored with explanatory variables such as grade level, gender, and student attitude toward writing? To answer these questions, our analytic approach relied on invariant measurement. Invariant measurement (Engelhard 2013) draws on Rasch measurement theory (Rasch 1960/1980), and this framework is used to investigate the psychometric quality of the MPAW Scale. Invariant measurement is based on the requirement that instruments, including their meaning and use, remain consistent across different subgroups of students. If we create a stable and invariant frame of reference, then we begin to consider substantive differences in student perspectives within and between students. The Facets computer program was used to produce Wright maps that visually depict the relationships among persons (students), items, and other variables, as well as model-data fit statistics that provide support for inferences regarding how well the Wright map represents the latent variable of perceived authenticity (Linacre 1989).

For research questions one and two regarding item difficulties and model-data fit, we created Model 1 that included only persons and items. For research question three regarding explanatory variables, we created Model 2 that included persons, items, and three explanatory variables: student attitude toward writing, gender, and grade level. Both models used the partial credit model and the equations for each model are presented below:

Model 1

$$\ln \frac{P_{nijk}}{P_{nijk-1}} = \theta_n - \delta_i - \tau_{ik}$$

Model 2

$$\ln \frac{P_{nijk}}{P_{nijk-1}} = \theta_n - \delta_i - \Delta_j - \tau_{ik}$$

where

P_{nijk} the probability of student n responding in category k on item i;
P_{nijk-1} the probability of student n responding in category $k-1$ on item i;
θ_n the perception of authenticity by student n;
δ_i the location of item i;
Δ_j the location of explanatory variable j; and
τ_{ik} the difficulty of responding in category k relative to $k-1$ on item i.

The explanatory variables, Δ_j, included in this study are grade, gender and attitude toward writing.

Instrument

The Modified Perceived Authenticity in Writing (MPAW) Scale consists of 16 items with a 6-point Likert Scale (1-Strongly Disagree to 6-Strongly Agree) (see Appendix for details). One original item from the PAW Scale was dropped, and language was changed from task-specific to general for all other items. For example, an original item on the PAW Scale states, "Writing this paper helped me to understand the topic better" and the corresponding item on the MPAW Scale states, "Writing in my English language arts class helps me to understand topics better." In addition to responding to the 16 items, students also indicated their grade level (6–9), their interest in writing (from 1–6 with 1 indicating the lowest interest and 6 representing the highest interest), and their gender (Male, Female, or Other.) These demographic questions were used to explore how the scale may function differently for different groups of students.

Participants

Seventy-four students at one school site completed the MPAW Scale during an after-school program in the spring of 2015. All students provided written assent, and they had written parental consent to participate. Students mostly identified as Black or African American, 99 % participated in free and reduced school lunch programs, and their ages ranged from 11- to 14-years old.

Data Collection and Data Analysis

The MPAW Scale was administered in paper format to the 74 students in the study by a research assistant. Demographic questions preceded the MPAW Scale items. The research assistant first introduced the purpose of the study, indicating that our goal was to understand students' views on their current writing instruction in their English language arts class. Then the research assistant read through the demographic questions and instructed students to choose the responses that best represented them. Next, the research assistant read through the instructions for the MPAW Scale and then read the items, pausing after each item so students could record their answers. There were also additional Likert items and two short answer questions students answered in addition to the MPAW Scale items, but these items are not analyzed in the current study.

After data were collected, all data were entered into an Excel spreadsheet and then exported to the Facets program to run the Rasch analyses, including conversion of the raw ordinal data into interval data, creation of person and item fit statistics, and creation of Wright maps that visually display person and item locations and any additional variables included in particular models. For our analyses, the original 6-point Likert scale structure was maintained and ratings were not collapsed. Our findings based on these analyses are described in detail in the next section.

Findings

Model 1 Results

The first analysis explores if the internal structure of the MPAW Scale represents gradations in item difficulty and if items exhibit acceptable model-data fit. Overall, the Rasch model explained 51.7 % of the variance of the MPAW Scale. The results of the Rasch model indicated that the MPAW Scale has reasonably high reliability of person separation with $Rel_{Students} = 0.89$. Additionally, the scale has a relatively high reliability of item separation statistic with $Rel_{Items} = 0.77$. Table 1 contains the summary statistics for Model 1. These relatively high reliability statistics indicate that the MPAW Scale's items can be separated from one another and can be used to differentiate students' perceptions of authenticity, suggesting that the internal structure of the scale does indeed represent gradations of item difficulty.

Figure 1 contains the Wright map for Model 1, which displays the spread of persons and items graphically. This figure visually presents the item and person

Table 1 Summary statistics for Model 1

	Students	Items
Mean	0.32	0.00
SD	1.33	0.22
N	74	16
Infit M	1.11	0.99
Infit SD	0.93	0.31
Outfit M	1.16	1.16
Outfit SD	1.17	0.67
Reliability of separation	0.89	0.77
Chi-square	606.0[*]	68.7[*]
Df	73	15

[*]p < 0.01

separation statistics noted above. Column 1 is the scale in logits that acts as a common *ruler* in Rasch measurement theory to examine the relationship between persons and items. The next two columns present the locations of persons and items on the logistic scale. Higher items on this scale indicate lower levels of endorsement, meaning the item was more "difficult" to endorse, whereas lower locations for items indicate higher levels of endorsement.

Thus, Item 6, "I discuss the topics of my ELA writing assignments with my family," was the most difficult item to endorse, and Item 11, "I am proud of what I write in my ELA class," was the easiest item to endorse. Based on the trend in the literature of school writing tending to be a classroom contained activity versus a broader activity connecting to the community, these levels of endorsement make theoretical sense.

The person separation statistics, item separation statistics, and Wright map indicate that there is a hierarchy of item difficulty. Moving to the second research question regarding model-data fit, Wright and Linacre (1994) suggest that acceptable indices of model-data fit are obtained when the Infit and Outfit statistics range from 0.60 to 1.40. Table 2 presents the item quality index for all items in the scale, and Infit and Outfit statistics. As can be seen in Table 2, there are five items exhibiting some misfit. Item 1 is the only item with both Infit and Outfit statistics outside the acceptable range, while Items 2, 3, and 4 had Outfit statistics outside the acceptable range. Item 12 had a generally higher than acceptable Infit statistic.

These misfitting items suggest that there may not be a consistent difficulty hierarchy for these items. This makes sense when thinking about authenticity as a subjective judgment because students may value particular factors of authenticity above others and additionally, students may perceive their writing instruction differently. We return to the misfitting items in the discussion section. We also propose modifications that may improve model-data fit.
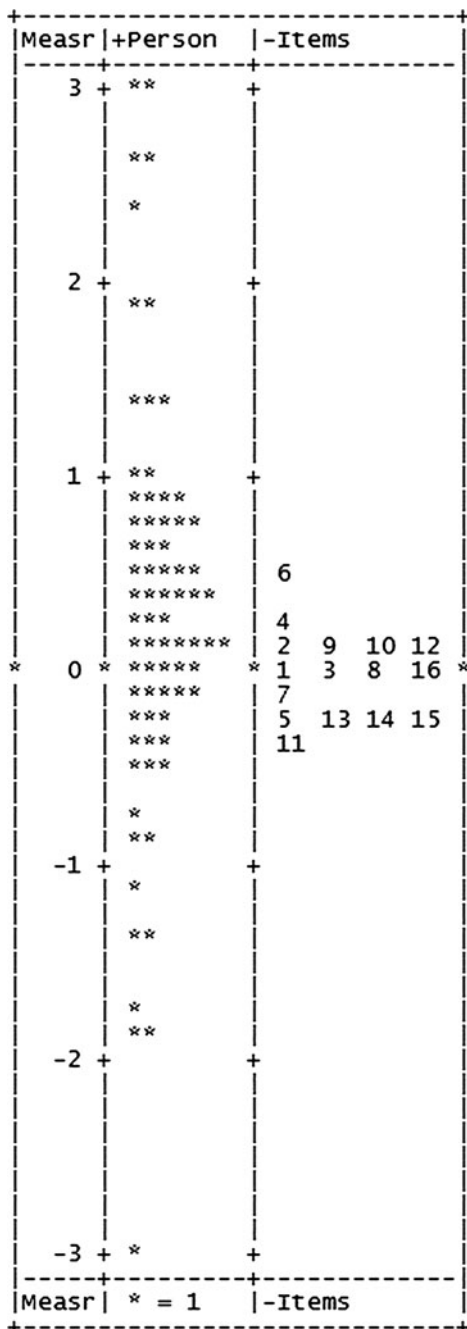
```
+-------------------------------+
|Measr |+Person  |-Items        |
|------+---------+--------------|
|   3 + **       +              |
|      |                        |
|      |                        |
|      |  **     |              |
|      |                        |
|      |  *      |              |
|      |                        |
|      |         |              |
|   2 +          +              |
|      |  **     |              |
|      |                        |
|      |                        |
|      |  ***    |              |
|      |                        |
|      |                        |
|   1 +  **      +              |
|      |  ****   |              |
|      |  *****  |              |
|      |  ***    |              |
|      |  *****  | 6            |
|      |  ****** |              |
|      |  ***    | 4            |
|      |  *******| 2   9   10 12 |
| *  0 *  *****  * 1   3   8   16 *
|      |  *****  | 7            |
|      |  ***    | 5   13 14 15 |
|      |  ***    | 11           |
|      |  ***    |              |
|      |                        |
|      |  *      |              |
|      |  **     |              |
|  -1 +          +              |
|      |  *      |              |
|      |                        |
|      |  **     |              |
|      |                        |
|      |  *      |              |
|      |  **     |              |
|  -2 +          +              |
|      |                        |
|      |                        |
|      |                        |
|      |                        |
|      |                        |
|  -3 +  *       +              |
|------+---------+--------------|
|Measr | * = 1   |-Items        |
+-------------------------------+
```

Fig. 1 Wright map for Model 1

Table 2 Item quality index in Rasch analysis

Items	Mean Rating	Measure	SE	Infit MS	Outfit MS
1. The writing that I do in my ELA class is related to my life outside of class	3.88	0.06	0.12	1.93[*]	2.38[*]
2. I enjoy writing in my ELA class	3.84	0.13	0.10	0.64	0.56[*]
3. ELA writing assignments relate to topics I care about in the world	3.85	0.05	0.10	1.14	2.29[*]
4. People other than my teacher read the papers I write for school	3.62	0.25	0.09	1.39	2.20[*]
5. I will use what I am learning about writing to write other papers in the future	4.24	−0.21	0.10	0.82	0.82
6. I discuss the topics of my ELA writing assignments with my family	3.42	0.48	0.10	0.88	0.85
7. What I am learning about writing is important for my life	4.08	−0.11	0.11	0.89	0.87
8. ELA assignments are important to me	4.00	0.00	0.1	0.82	0.74
9. Writing in my ELA class connects to my personal interests	3.79	0.17	0.10	0.92	1.12
10. People who read my ELA writing assignments will change their opinions, actions, or feelings	3.91	0.11	0.10	0.92	0.91
11. I am proud of what I write in my ELA class	4.36	−0.38	0.12	1.13	1.04
12. I discuss the topics of my writing assignments with friends	3.81	0.18	0.10	1.80[*]	1.04
13. I am gaining writing skills that I will use later in my life in my ELA class	4.34	−0.23	0.11	0.89	0.79
14. Writing in my ELA class helps me develop my thoughts, opinions, or beliefs	4.36	−0.25	0.12	0.78	0.81
15. Writing in my ELA class is making me a better writer	4.26	−0.24	0.11	0.73	0.78
16. Writing in my ELA class helps me to understand topics better	4.00	0.00	0.11	0.81	0.78

Notes (1) English language arts is abbreviated to ELA in this Table
(2) [*]Indicates Infit and Outfit statistics indicating item misfit

Model 2 Results

The second model included items and persons, and also added three explanatory variables. This model was used to answer the third research question: Do the MPAW items exhibit measurement invariance when explored with explanatory variables such as grade level, gender, and student attitude toward writing? Based on our analyses, differences on MPAW Scale by all explanatory factors were

statistically significant at a 0.01 level. This means that in this study, perceptions of authenticity varied significantly by grade level, by gender, and by student attitude toward writing. This finding aligns with past research (Behizadeh 2014, 2015) that found that authenticity varied by student characteristics, including gender and ethnicity. Potentially due to shared characteristics of a subgroup (e.g., cultural background), particular subgroups within a student population may perceive the authenticity of writing instruction at different levels.

Figure 2 is the Wright map for Model 2, and it graphically displays the relationships between items, students, and explanatory variables. As can be seen in this graphic representation, students who had high interest in writing (represented by a 6 on the 6-point scale) also were more likely to have higher scores on the MPAW Scale. This connection is theoretically logical since both authenticity and writing interest are motivational variables and a highly motivated student who has a high interest in writing may perceive instruction as more authentic and connected to their life. Additionally, males and those who indicated "Other" for gender were more likely to have higher scores on the MPAW Scale than females, a difference that does not have a clear explanation based on the literature. Also, in terms of grade level, those in the 6th grade were more likely to have higher scores on the MPAW Scale while those in the 8th grade were more likely to have lower scores. As a possible explanation for this difference, students in different grades are experiencing different curricula, which may include features that raise or lower perceptions of authenticity; yet this hypothesis cannot be confirmed by the data collected in the current study. Although there already exists a strong theoretical rationale for the connection between higher scores on the MPAW Scale and higher interest in writing, the underlying reasons for males and 6th graders having higher scores would need to be confirmed in future studies drawing on larger sample sizes and also explored further with qualitative methods.

Regarding interactions between items and explanatory variables, there were no significant interactions between items and student attitude toward writing, grade level, or gender. Thus, the MPAW items exhibit measurement invariance when explored with the explanatory variables of grade level, gender, and student attitude toward writing.

Discussion

Returning to research questions 1 and 2, the Wright map for Model 1 suggests a hierarchy of item difficulties, and there was high reliability of item and person separation. Regarding research question 3, the MPAW items exhibit measurement invariance when explored with the explanatory variables of grade level, gender, and student attitude toward writing. These findings suggest that the scale is able to differentiate students in terms of perceived authenticity, and that the internal structure of the MPAW is stable. Using the principles of invariant measurement, we identified several misfitting items. These items may not be consistently interpreted
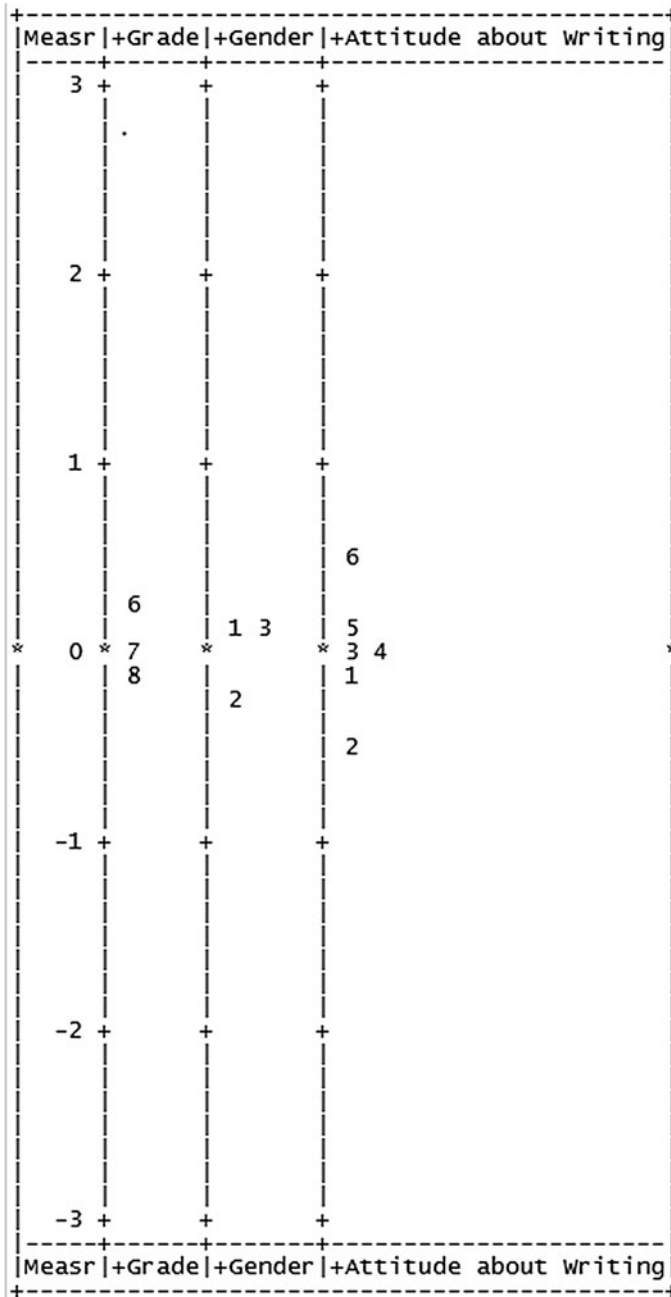
```
+-------------------------------------------------+
|Measr |+Grade |+Gender |+Attitude about writing|
|------+-------+--------+-----------------------|
|   3 +       +        +                         |
|      |   .   |        |                        |
|      |       |        |                        |
|      |       |        |                        |
|      |       |        |                        |
|   2 +       +        +                          |
|      |       |        |                        |
|      |       |        |                        |
|      |       |        |                        |
|      |       |        |                        |
|   1 +       +        +                          |
|      |       |        |                        |
|      |       |        |   6                     |
|      |   6   |        |                         |
|      |       |  1 3   |   5                     |
*  0 * 7     *        * 3 4                        *
|      |   8   |        |   1                      |
|      |       |   2    |                          |
|      |       |        |   2                      |
|      |       |        |                          |
|  -1 +       +        +                           |
|      |       |        |                         |
|      |       |        |                         |
|      |       |        |                         |
|      |       |        |                         |
|  -2 +       +        +                           |
|      |       |        |                         |
|      |       |        |                         |
|      |       |        |                         |
|      |       |        |                         |
|  -3 +       +        +                           |
|------+-------+--------+-----------------------|
|Measr |+Grade |+Gender |+Attitude about writing|
+-------------------------------------------------+
```

Fig. 2 Variable map for Model 2 [*Notes* Gender (1 = Male; 2 = Female; 3 = Other); Attitude (1 = low interest; 6 = high interest)]

in terms of a hierarchal structure for all items. However, unlike with the measurement of achievement in certain content areas (such as math) where there should be a more or less orderly progression of difficulty based on students levels of knowledge, the measurement of affective variables is more subjective; it may be that some students value certain features of authenticity more than other students. Additionally, because the questions are around how students perceive authenticity, individual students may perceive the same writing instruction differently.

However, revisions to the MPAW Scale could result in better model-data fit. For example, looking at the data, we found that students are not using all categories; the lower end of the scale was underutilized in this study, and other researchers have condensed scales that reveal this underutilization of categories (e.g., Engelhard and Chang 2015). Although it could be that students in different contexts that contain less features of authentic writing instruction will use the full scale, we wondered if perhaps there are not as many gradations for perceived authenticity as six. Future research should experiment with using a three-point scale and investigate if this structure improves model-data fit.

As another possible solution some items may be dropped or re-written. For example, Item 2, "I enjoy writing in my ELA class," could be misfitting because it is measuring enjoyment versus authenticity. Potentially, students can rate classroom writing instruction as highly authentic yet not enjoy it. This item may be dropped in future administrations. Also, Item 1, "The writing that I do in my ELA class is related to my life outside of class" may be endorsed by students who are perceiving the authenticity of their classroom writing instruction at very different levels, as long as there is some connection to their lives outside of the classroom. This could be rewritten as "The writing that I do I my ELA class is strongly connected to my life outside of class," which may make this item harder to endorse for those in low to medium authenticity classrooms and thus yield better model-data fit.

Finally, three items exhibiting some misfit may need to be revised: Item 3, "ELA writing assignments relate to topics I care about in the world;" Item 4, "People other than my teacher read the papers I write for school;" and Item 12, "I discuss the topics of my writing assignments with friends." The items identify features that may increase authenticity for some but not for all students. Thinking about the subjective nature of authenticity, certain students may value external readers (Item 4) when considering the overall authenticity of a task, while other students may not value this element of writing instruction. Similarly, discussing topics of writing assignments with friends (Item 12) or writing about issues of global import (Item 3) may matter more for some students than others. For these three items that exhibit misfit, future qualitative research with students investigating factors of authenticity may help refine the language so that these specific factors can represent broader, more universal features of authenticity. Thus, a major recommendation moving forward in addition to the minor modifications suggested here is to pair student and teacher interviews with MPAW Scale use. Because of the complexity of the construct and the lack of research on measuring authenticity, qualitative data can be used to support the future revisions of the MPAW Scale and interpretation of MPAW Scale use.

Conclusion

Based on our analyses, we believe the MPAW Scale is potentially a useful tool for examining overall impressions of authenticity of writing. This study provides evidence that the scale is reliable, as well as some validity evidence for use with students of color in an urban setting. Although the participants in the current study did not represent the full range of ethnicity and socioeconomic status in the United States, one of our major goals in our program of research is to examine perceived authenticity for historically underserved students, and the piloting of the scale with this particular subgroup was a strategic decision. However, future research is needed in different contexts to examine if the scale operates differently (e.g., differential item functioning) for different subgroups of students.

These analyses serve as the base for future studies that will examine teacher and student perspectives on writing instruction and assessment. Because consequential validity is a key facet of a holistic view of validity, an instrument that can easily and accurately capture student perceptions of the authenticity of writing instruction can be a useful source of validity evidence, especially when paired with qualitative data to support interpretation of quantitative data. Future work will include qualitative data sources, such as student and teacher interviews to help interpret authenticity data. If large-scale writing assessments are linked to writing instruction characterized by low authenticity on scales such as the MPAW Scale, these assessments may need to be revised.

Honoring and prioritizing consequences aligns with a vision of writing assessment research that considers students as primary stakeholders in the assessment process (Behizadeh and Engelhard 2014; Guba and Lincoln 1989; Slomp et al. 2014) who should be protected from outcomes (regardless of intention) that are damaging to students' affective or cognitive development, as well as their academic achievement. Researchers have often determined what counts as authentic for students, rather than asking students themselves what they need for authentic education and authentic writing instruction. Students are important stakeholders in large-scale assessments, and their perspectives are underrepresented in discussions of reliability, validity, and fairness of score meaning and use. Soliciting student perspectives on authenticity or other affective variables during the validation process will offer another source of validity evidence that can be used to examine consequential validity, such as the access of historically underserved students to engaging, authentic writing instruction.

Appendix

Comparison of PAW Scale and MPAW Scale

PAW Scale	Modified PAW Scale
1. This writing assignment was relevant and/or meaningful to my life outside of class	1. The writing I do in my English language arts class is related to my life outside of class
2. People other than my teacher will want to read the paper I wrote	4. People other than my English teacher read the papers I write for school
3. Writing this paper was a good learning experience	15. Writing in my English language arts class is making me a better writer
4. I can make connections between this paper and events or issues in the world that I care about	3. English language arts writing assignments relate to topics I care about in the world
6. I will use what I learned writing this paper to write other papers	5. I will use what I am learning about writing to write other papers in the future
7. I have discussed or will discuss the topic of this paper with family members	6. I discuss the topics of my English language arts writing assignments with my family
8. I enjoyed writing this paper	2. I enjoy writing in my English language arts class
9. I think knowing how to write a paper like this one will be important to know in my life	7. What I am learning about writing is important to know in my life
10. Writing this paper was important to me	8. English language arts writing assignments are important to me
11. This paper connects to my personal interests	9. Writing in my English language arts class connects to my personal interests
12. People who read this paper will change their opinions, actions, or feelings	10. People who read my English language arts writing assignments will change their opinions, actions, or feelings
13. I am proud of what I wrote	11. I am proud of what I write in my English language arts class
14. Writing this paper helped me to understand the topic better	16. Writing in my English language arts class helps me to understand topics better
15. I have discussed or will discuss the topic of this paper with friends	12. I discuss the topics of my writing assignments with friends
16. I will use the skills that I learned writing this paper later in my life	13. I am gaining writing skills that I will use later in my life in my English language arts class
17. Writing this paper helped me to develop my thoughts, opinions, or beliefs	14. Writing in my English language arts class helps me develop my thoughts, opinions, or beliefs
5. This paper connected to something I recently saw on TV or the internet[*]	

[*]This item is not included in the modified scale

References

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal, 100*(6), 14–27.

Ashton, S. (2010). Authenticity in adult learning. *International Journal of Lifelong Education, 29*(1), 3–19.

Au, W., & Gourd, K. (2013). Asinine assessment: Why high-stakes testing is bad for everyone, including English teachers. *English Journal, 103*(1), 14–19.

Ball, A. F., & Ellis, P. (2008). Identity and the writing of culturally and linguistically diverse students. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 499–513). New York, NY: Lawrence Erlbaum.

Behizadeh, N. (2014). Adolescent perspectives on authentic writing. *Journal of Language and Literacy Education, 10*(1), 27–44. Retrieved from http://jolle.coe.uga.edu.

Behizadeh, N. (2015). Engaging students through authentic and effective literacy instruction. *Voices from the Middle, 23*(1), 40–50.

Behizadeh, N., & Engelhard, G. (2014). Development and validation of a scale to measure perceived authenticity in writing. *Assessing Writing, 21*, 18–26.

Common Core State Standards Initiative. (2015). *English language arts standards.* Retrieved from http://www.corestandards.org/ELA-Literacy/.

Council of Chief State School Officers. (2011). *InTASC model core standards at a glance.* Retrieved from http://www.ccsso.org/Resources/Publications/InTASC_Standards_At_a_Glance_2011.html.

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.

Dyson, A. H., & Freedman, S. W. (2003). Writing. In J. Flood, D. Lapp, J. R. Squire, & J. M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 967–992). Mahwah, NJ: Lawrence Erlbaum.

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

Engelhard, G., & Chang, M. (2015). Examining the teachers' sense of efficacy scale at the item level with Rasch measurement model. *Journal of Psychoeducational Assessment*, 1–15.

Fisher, M. T. (2007). *Writing in rhythm: Spoken word poetry in urban classrooms*. New York, NY: Teachers College Press.

Freire, P. (1970/2000). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). New York, NY: Continuum. (Original work published 1970).

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Haertel, E. H., Moss, P. A., Pullin, D. C., & Gee, J. P. (2008). Introduction. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 1–16). Cambridge: Cambridge University Press.

Hillocks, G, Jr. (2011). Commentary on "Research in secondary English, 1912–2011: Historical continuities and discontinuities in the NCTE imprint". *Research in the Teaching of English, 46*(2), 187–192.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

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

Luna, C., & Turner, C. L. (2001). The impact of the MCAS: Teachers talk about high-stakes testing. *English Journal, 91*(1), 79–87.

Madaus, G. F. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's testing policy. *Harvard Educational Review, 64*(1), 76–95.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Morrell, E. (2008). *Critical literacy and urban youth: Pedagogies of access, dissent, and liberation*. New York, NY: Routledge.

National Council of Teachers of English and International Reading Association. (2012). *Standards for the English language arts*. Retrieved from http://www.ncte.org/standards/ncte-ira.

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*(4), 280–312.

Partnership for 21st Century Learning. (n.d.) Framework for 21st century learning. Retrieved from http://www.p21.org/our-work/p21-framework.

Purcell-Gates, V., Anderson, J., Gagne, M., Jang, K., Lenters, K. A., & McTavish, M. (2012). Measuring situated literacy activity: Challenges and promises. *Journal of Literacy Research, 44*(4), 396–425.

Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly, 42*(1), 8–45.

Sisserson, K., Manning, C. K., Knepler, A., & Jolliffe, D. A. (2002). Authentic intellectual achievement in writing. *English Journal, 91*(6), 63–69.

Seunarinesingh, K. (2010). Primary teachers' explorations of authentic texts in Trinidad and Tobago. *Journal of Language and Literacy Education [Online], 6*(1), 40–57.

Slomp, D. H., Corrigan, J. A., & Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments: A Canadian study. *Research in the Teaching of English, 48*(3), 276–302.

Splitter, L. J. (2009). Authenticity and constructivism in education. *Studies in Philosophy and Education, 28*, 135–151.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press. (Original work published 1960).

Watanabe, M. (2007). Displaced teaching and state priorities in a high-stakes accountability context. *Educational Policy, 21*(2), 311–368.

Winn, M. T., & Johnson, L. (2011). *Writing instruction in the culturally relevant classroom*. Urbana, IL: National Council of Teachers of English.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch measurement transactions, 8*(3), 370.

Author Biographies

Dr. Nadia Behizadeh is an Assistant professor of adolescent literacy at Georgia State University in the United States. Her research explores authentic and culturally sustaining literacy instruction and assessment for adolescent learners, with a focus on increasing the impact of adolescent student writing.

Dr. George Engelhard, Jr. is a professor of educational measurement and policy at The University of Georgia in the United States. He is a fellow of the American Educational Research Association. His latest book is Invariant measurement: Using Rasch models in the social, behavioral, and health sciences (New York: Routledge).