



## OPEN ACCESS

## EDITED BY

Ariel Mariah Lindorff,  
University of Oxford, United Kingdom

## REVIEWED BY

Mustafa Erol,  
Yildiz Technical University, Türkiye  
Argyris Nipyrakis,  
University of Crete, Greece

## \*CORRESPONDENCE

Angela Crawford  
✉ angelacrawford1@boisestate.edu

RECEIVED 10 March 2025

ACCEPTED 07 July 2025

PUBLISHED 30 July 2025

## CITATION

Crawford A and Starrett A (2025) Bridging contexts in STEM professional development: validating a social validity measure across rural and nonrural educators.  
*Front. Educ.* 10:1591267.  
doi: 10.3389/feduc.2025.1591267

## COPYRIGHT

© 2025 Crawford and Starrett. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Bridging contexts in STEM professional development: validating a social validity measure across rural and nonrural educators

Angela Crawford<sup>1\*</sup> and Angela Starrett<sup>2</sup>

<sup>1</sup>Department of Mathematics, Boise State University, Boise, ID, United States, <sup>2</sup>Department of Educational and Developmental Science, University of South Carolina, Columbia, SC, United States

**Introduction:** This study investigates the measurement invariance of a survey assessing perceptions of STEM professional development (PD) resources among rural and non-rural educators. Social validity theory provides the framework for examining four constructs: feasibility, usability, appropriateness, and local relevance.

**Methods:** Using a mixed-methods approach, we conducted multi-group confirmatory factor analysis to test for configural, metric, scalar, and strict invariance. We also conducted a qualitative thematic analysis of open-ended responses to provide contextual insights and nuance.

**Results:** Results indicate full measurement invariance, supporting the validity of cross-group comparisons. Quantitative analyses show that rural educators rated PD resources significantly higher in feasibility and appropriateness, while no significant differences emerged for usability or local relevance. Thematic analysis revealed that rural teachers more frequently emphasized professional learning benefits and expressed more positive sentiment toward PD resources.

**Discussion:** These findings highlight the importance of designing STEM PD initiatives that account for contextual differences in resource accessibility, instructional autonomy, and community relevance. Implications for policy and practice include recommendations for tailoring PD to diverse educational settings to enhance educator engagement.

## KEYWORDS

measurement invariance, STEM professional development, social validity, rural education, mixed-methods research, survey validation, educational contexts

## Introduction

In the evolving landscape of education, professional development (PD) of STEM educators is recognized as a critical mechanism for enhancing teaching practices, fostering inquiry-based learning, and ultimately improving student achievement in STEM disciplines. However, the effectiveness and uptake of PD initiatives often differ significantly across varying educational contexts, particularly between rural and non-rural settings (Desimone, 2009; Penuel et al., 2007). Rural educators face unique challenges, including limited access to resources, professional isolation, and disparities in funding (Howley and Howley, 2005; Harmon, 2020). Addressing these challenges requires understanding how PD resources are perceived by educators across different environments, which necessitates examining the measurement tools used to assess these perceptions.

The STEM PD Resource Use Survey was developed to explore teachers' perceptions of feasibility, usability, appropriateness, and local relevance of PD resources, drawing upon social validity theory (Wolf, 1978). Social validity theory emphasizes three key components: socially significant goals, acceptable processes, and important effects. By assessing teachers' perceptions, PD developers and educational researchers can better align resources with teachers' needs, thereby increasing the likelihood of effective implementation.

The constructs assessed by the survey (i.e., feasibility, usability, appropriateness, and local relevance) are key factors influencing teachers' adoption of PD resources. Feasibility refers to contextual support and barriers to implementation, while usability addresses a teacher's assessment of the resource's clarity and comprehensiveness (Greer et al., 2012). Appropriateness examines whether the resource provides sufficient benefits to justify its use, and local relevance considers the degree to which PD resources are aligned with community-specific needs, such as vocational opportunities for students (Howard et al., 2024). These constructs provide a nuanced understanding of the factors that may influence PD adoption across different educational contexts.

Given the distinct educational environments of rural and non-rural schools, it is vital to confirm that any observed differences in survey responses genuinely reflect differences in perceptions rather than biases or inconsistencies in the measurement tool itself (Rutkowski and Svetina, 2014). Measurement invariance is a fundamental requirement for comparing survey scores across groups; without it, differences in scores could be erroneously attributed to the context rather than to true differences in perception or experience (Villarreal-Zegarra et al., 2019).

This study uses a mixed-methods approach to examine rural and non-rural educators' perceptions of STEM PD resources. The quantitative component focuses on assessing measurement invariance through multi-group confirmatory factor analysis (MG-CFA), while the qualitative component involves thematic analysis of open-ended survey responses to provide deeper insights into the context-specific nuances of PD resource use. This approach allows us to address the broader challenge of bridging the uniqueness of rural education contexts with the universality of effective PD practices. By integrating quantitative and qualitative perspectives, this study aims to contribute to the growing body of research on the social validity of PD resources, offering evidence for their applicability across diverse educational settings and providing insights that can inform the design of more context-sensitive PD initiatives.

## Theoretical framework

Social validity theory is concerned with what it means for something to be viewed as socially important or valuable (Wolf, 1978). Social validity applies three conceptual components to evaluation of a program or resources: socially significant goals, acceptable processes, and importance of effects. Gathering information about goals, processes, and effects can help researchers understand why practices may or may not be adopted and can help PD developers improve PD resources.

Goals, processes, and effects are reflected in scales in various ways depending upon the purpose of the scale and the context in which it is applied (e.g., Greer et al., 2012; Lakin and Shannon, 2015; Weiner et al., 2017). A scale may assess value based on practical or personal criteria (Weiner et al., 2017). Additionally, assessments may be influenced by features of the broad educational context (an external influence) or characteristics of teachers and their classroom instruction (an internal influence; Weiner et al., 2017). We identified four constructs related to goals, processes, and effects in teachers' decisions to implement PD resources that reflect these criteria and influences: feasibility, usability, appropriateness, and local relevance. These four constructs were selected because they directly align with Wolf's (1978) framework: feasibility and usability reflect the *acceptability of processes*, appropriateness indicates the *importance of effects*, and local relevance is tied to the *social significance of goals* within a specific geographic and cultural context.

These constructs also represent how teachers express their professional development values in practice. Teacher professional development values refer to the beliefs, judgments, and contextual preferences that shape educators' decisions about whether and how to engage with PD opportunities. These values influence not only how teachers interpret the goals and utility of PD resources but also whether they perceive those resources as implementable or worthwhile within their specific classroom or community settings (Weiner et al., 2017; Biddle and Azano, 2016). For rural educators, PD values are often shaped by resource constraints, professional isolation, and the imperative to connect instruction with local relevance or community needs (Howley and Howley, 2005; Harmon, 2020).

In this study, we operationalize teacher PD values using four constructs rooted in social validity theory: feasibility, usability, appropriateness, and local relevance. These constructs were selected to represent key aspects of teachers' values regarding PD implementation. Feasibility and usability reflect the acceptability and practicality of PD processes, appropriateness captures the perceived value or worth of engaging with the resource, and local relevance aligns with teachers' judgments about the social or geographic meaningfulness of the content. This approach is consistent with prior implementation research emphasizing the importance of personal, contextual, and community-based criteria in how teachers evaluate interventions (Greer et al., 2012; Weiner et al., 2017).

## Feasibility

Feasibility is concerned with contextual support and barriers to implementation, and whether the change in practice can be performed given the resources available to the teacher. Feasibility of the implementation of PD resources may be related to access to additional necessary resources, administrator support, time needed for implementation, and costs.

## Usability

Usability refers to a teacher's assessment of the clarity and comprehensiveness of the resources and necessary guidance for

implementation. While sharing a practical focus with feasibility, usability is a personal assessment. Two individuals in the same context may have different assessments of the usability, possibly due to factors such as background knowledge or teaching style. Thus, usability addresses questions about PD resources' potential for supporting implementation by a range of individual users.

## Appropriateness

We define appropriateness as providing advantages or benefits which make efforts toward implementation worthwhile. It considers whether PD resources are worthy of the time and effort expended for a teacher to learn to use the resources and make changes to their regular classroom practice. Appropriateness is concerned with personal choice when faced with competing instructional opportunities.

## Local relevance

Education is situated in place – geographic, economic, and/or cultural spaces. These places affect what is considered a significant goal and a meaningful outcome. Teachers working in the same community are more likely to identify similar place-based goals and meaningful outcomes. Local relevance includes practical and social applicability, pertinence, and connections to the community. Local relevance considers the resources' value for positive impact on students' lives in terms of geographic, cultural, societal, and vocational dimensions (Howard et al., 2024; O'Neill et al., 2023; Stuckey et al., 2013; Xu et al., 2023).

## Research questions

While social validity is widely acknowledged as essential for evaluating educational interventions, few studies examine whether perceptions of PD resources are comparably structured and interpreted across distinct educational contexts. Rural educators, often faced with limited access to high-quality PD and contextually misaligned resources, remain underrepresented in implementation and measurement research (Biddle and Azano, 2016; Harmon, 2020; Player, 2015). Without validating instruments across groups, researchers risk misattributing group differences to context rather than measurement artifacts (Rutkowski and Svetina, 2014).

This study addresses these gaps by using a mixed-methods approach to examine whether an instrument assessing teachers' perceptions of PD resources operates consistently across the distinct environments of rural and non-rural STEM education. We operationalize teachers' professional development values through four constructs (i.e., feasibility, usability, appropriateness, and local relevance), each aligned with components of social validity theory. Specifically, we test whether the survey demonstrates measurement invariance, compare latent means, and explore open-ended responses to understand the contextual factors shaping PD perceptions. These objectives guide the following research questions:

1. To what extent does the survey measuring feasibility, usability, appropriateness, and local relevance exhibit configural, metric,

scalar, and strict invariance between rural and non-rural STEM teachers?

2. Among the constructs (feasibility, usability, appropriateness, local relevance), do rural and non-rural STEM teachers differ in their latent mean scores, and if so, which constructs show the most substantial differences?
3. What do open-ended responses reveal about the contextual factors (e.g., resource availability, community values, administrative support) that might explain why certain aspects of STEM PD resources resonate differently among rural and non-rural educators?

## Methods

A convergent parallel mixed-methods design (Creswell and Plano Clark, 2017) was used to investigate the differences and similarities in responses of rural and non-rural teachers to the STEM PD Resource Use Survey. In this design, quantitative and qualitative data are collected simultaneously, analyzed separately, and then integrated during interpretation to provide a comprehensive understanding of the research problem. The quantitative strand involved multi-group confirmatory factor analysis (MG-CFA) to assess measurement invariance across groups, while the qualitative strand used thematic analysis to examine teachers' open-ended comments about PD resource use.

## Survey

The STEM PD Resource Use Survey was developed using an argument-based validation approach (American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), 2014) to assess teachers' perceptions of PD resources. The current study provides evidence for the claim that the instrument exhibits an internal structure that is consistent across two relevant populations—rural and non-rural STEM teachers. The survey design followed best practices in measure development (Holmbeck and Devine, 2009), beginning with an extensive literature review, expert consultations, and cognitive interviews with teachers. This process ensured content validity by aligning the survey constructs with social validity theory (Wolf, 1978). Items were refined through cognitive interviews with teachers and feedback from PD facilitators to ensure alignment with the targeted constructs. Based on participant feedback, item modifications, additions, and deletions were made to improve clarity and applicability. Each item was evaluated with a 5-point Likert scale (strongly disagree, disagree, neutral, agree, strongly agree). The final instrument contained 26 Likert-type items, later refined to 16 items based on psychometric evaluation (items are listed in Table 1).

A confirmatory factor analysis (CFA) was conducted to validate the factor structure of the survey using a sample of teachers ( $n = 398$ ) who had participated in STEM PD workshops. Given the ordinal nature of the Likert scale responses, robust maximum likelihood estimation (MLR) was used in lavaan (Rosseel et al., 2023). After refining to 16 items, the final model demonstrated acceptable global fit ( $\chi^2(97) = 231.89, p < 0.001$ ; CFI = 0.949,

TABLE 1 Descriptive statistics for sixteen social validity items in the STEM PD resource use survey.

Factor	Item	<i>n</i>	<i>M</i>	SD	skew	kurtosis
Feasibility	The content standards and grade level(s) I teach.	363	4.3	0.93	−1.47	1.77
	The learning needs of my students.	363	4.27	0.9	−1.39	1.78
	The time I have for planning.	363	3.99	1	−0.96	0.51
	The technology and other resources I have access to.	355	3.96	1.02	−1.02	0.64
	The priorities and initiatives in our school or district.	354	3.88	1.03	−0.78	0.06
Usability	Are easy to understand.	352	4.48	0.75	−1.77	3.78
	Are easy to implement.	351	4.2	0.95	−1.26	1.17
	Can be implemented in my context without many modifications.	352	3.88	1.1	−0.89	−0.06
	Provide enough support for me to implement in my classroom.	350	4.1	0.99	−1	0.41
Appropriateness	Are effective for teaching STEM content.	347	4.56	0.72	−1.98	4.87
	Fit easily into my instruction.	346	4	1.05	−0.99	0.36
	Are worth the effort it takes to learn and use them.	345	4.44	0.85	−1.69	2.67
Local relevance	Help me make connections between the local area and STEM topics.	374	4.12	0.91	−0.99	0.68
	Will benefit my students in the future.	373	4.51	0.76	−1.76	3.34
	Are a good match for the experiences my students have had in their lives.	374	4.13	0.9	−1.1	1.15
	Are relevant to jobs and industry in my area.	373	4.14	0.89	−0.95	0.55

Each question began with the phrase “The [PD program] PD resources...” or “The [PD program] PD resources are implementable given...”

RMSEA = 0.065, SRMR = 0.038). Factor loadings and inter-factor correlations further supported the four-factor structure. This validated factor structure serves as the foundation for the measurement invariance testing presented in this study.

## Participants and procedures

Participants were teachers who enrolled in STEM PD offered by a state agency in the previous 5 years. The state agency offers workshops annually in the summer. The goal is to develop teachers’ knowledge and pedagogical skills and support student interest in STEM. The workshops are offered to educators who work in classroom and informal settings. The workshops vary in content and are developed and facilitated by experts in STEM fields and/or STEM educators. The workshops take place in six locations across the state. Email addresses for the potential participants were provided by the state agency.

The survey was delivered via Qualtrics. The first page described the consent process as approved by an institution’s IRB. Using the email list provided, 1,212 invitations to participate were sent. We received a total of 398 responses (32.8%), including both complete and incomplete responses. In this sample, self-reported gender was 70% female, 14% male, 2% prefer not to say, and 14% provided no response. Sixty percent of respondents described their school or informal education setting as situated in a rural area, 39% described their setting as non-rural (urban or suburban), and 1% provided no response. Twenty-three respondents did not respond to the social validity Likert items on the survey. The proportion of missing data for variables ranged from 6% to 13%.

The analytic sample included 205 rural teachers and 129 non-rural teachers. While no *a priori* power analysis was conducted, simulation studies and general guidelines suggest that sample sizes of 200 or more per group are often adequate for testing

measurement invariance in models with a moderate number of indicators and parameters (Brown, 2015; Kline, 2023). Although the non-rural group was smaller ( $n = 129$ ), simulation studies indicate that invariance testing is robust to modest sample imbalance when the overall sample size exceeds 300 (Yoon and Lai, 2018). Thus, the analytic sample was considered adequate for the planned analyses.

## Quantitative analysis

To assess measurement invariance, MG-CFA was conducted separately for rural and non-rural groups. To assess CFA fit, we considered the root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), the comparative fit index (CFI), and the Tucker-Lewis index (TLI). RMSEA values  $\leq 0.08$  are considered to indicate acceptable fit and  $\leq 0.05$ –0.6 good fit (Bandalos, 2018). SRMR values of  $\leq 0.08$  are acceptable, and values  $\leq 0.05$  are considered good (Bandalos, 2018). CFI and TLI values  $\geq 0.95$  are indicative of good fit, while values  $> 0.90$  are acceptable (Bandalos, 2018). The four constructs—feasibility, usability, appropriateness, and local relevance—were modeled using confirmatory factor analysis, with multiple levels of invariance tested: configural, metric, and scalar. The lavaan package in R was used for analysis, utilizing maximum likelihood with robust standard error (MLR) estimation. Measurement invariance was examined to ensure that any observed differences in perceptions are not due to inconsistencies in how items function across the two groups.

The invariance assessment proceeded stepwise, beginning with configural invariance to establish whether the same factor structure is present across groups (rural and non-rural educators). Following this, metric invariance was tested to determine if factor loadings are equivalent across groups, indicating that the constructs



are perceived similarly. Then, scalar invariance was assessed to determine if item intercepts are equivalent, suggesting that the groups interpret the items similarly, allowing for meaningful comparisons of latent means. Lastly, residual invariance was explored to determine whether item-level measurement errors differ between rural and non-rural groups, ensuring that observed differences are due to substantive factors rather than differential response variability. Measurement invariance was evaluated based on conventional cutoffs:  $\Delta CFI \leq 0.01$ ,  $\Delta RMSEA \leq 0.015$ , and  $\Delta SRMR \leq 0.04$  between increasingly constrained models (Cheung and Rensvold, 2002). If these thresholds are exceeded, modifications would be explored to achieve partial invariance while maintaining comparability. If scalar invariance is confirmed, latent mean differences would be tested to compare group-level perceptions of PD resources.

## Qualitative analysis

The qualitative component involved a thematic analysis of open-ended responses from the survey. The thematic analysis was conducted to understand the context and nuances of teachers' responses related to PD resource use. Such an analysis can provide additional insights into the contextual factors that may influence teachers' responses and aid in interpreting any the results of the quantitative analysis.

Using NVivo 14®, responses to the open-ended question, “Is there anything you would like to share about the value or usefulness of the [PD program] resources?” were manually coded using inductive and deductive approaches. In the first phase, we used an inductive approach to create codes that describe the topic of each response. A constant comparison approach was used to identify codes for similar responses, and then codes were clustered under similar themes. Following this phase, participant responses were coded deductively for sentiment and alignment with the social validity framework. Codes for sentiment were very positive, moderately positive, moderately negative, and very negative (responses with no valence were not coded for sentiment). A priori codes for social validity were feasibility, usability, appropriateness, and local relevance. As we coded for social validity, we added two codes: access and general comment. Responses could be coded for more than one aspect of social validity as appropriate. Finally, each response was linked to whether the individual had identified the school as being situated in a rural or non-rural (urban or suburban) setting.

## Results

### Measurement invariance testing

We conducted a series of multi-group confirmatory factor analyses (MG-CFA) to assess measurement invariance for four constructs (i.e., local relevance, feasibility, usability, and appropriateness) across rural and non-rural STEM teachers. Following the guidelines of Cheung and Rensvold (2002) and Chen (2007), we compared increasingly constrained models (configural, metric, scalar, and residual/strict) and examined

changes in the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). While we report the Satorra-Bentler chi-square difference test, we refrain from using it to make decisions regarding invariance because it can be overly sensitive to sample size. Even trivial, practically unimportant misfits often yield a significant result, potentially leading to a Type 1 error (Cheung and Rensvold, 2002; Chen, 2007). Table 2 summarizes the fit statistics for each step.

The initial measurement model, assessed using MG-CFA, demonstrated acceptable fit indices for both rural and non-rural groups. The measurement model for the rural group showed acceptable fit with CFI = 0.95, TLI = 0.93, RMSEA = 0.065, and SRMR = 0.04. Similar values were observed for the non-rural group, with CFI = 0.94, TLI = 0.93, RMSEA = 0.08, and SRMR = 0.048, suggesting a good fit of the four-factor structure across both groups.

### Configural invariance

As shown in Table 2, the configural model, where only the pattern of factor loadings is the same across groups, had a CFI of 0.943, RMSEA of 0.071, and SRMR of 0.043. These values suggest acceptable fit, providing a baseline against which subsequent, more constrained models are compared.

### Metric invariance

Constraining factor loadings to be equal across rural and non-rural groups yielded a CFI of 0.936, RMSEA of 0.073, and SRMR of 0.086. These values suggest acceptable fit with SRMR just above the recommended cutoff of  $\leq 0.08$ . Additionally, SRMR is more sensitive to model complexity than CFI and RMSEA, which can often cause it to jump from configural to metric invariance. In following guidelines by Cheung and Rensvold (2002) and Chen (2007), we note the small changes in CFI and RMSEA as the primary tests for invariance. As such, the change in CFI relative to configural ( $-0.007$ ) remained below the commonly recommended cutoff of  $\leq |0.01|$  (Cheung and Rensvold, 2002), and the increase in RMSEA was only  $+0.001$ —well under the typical 0.015 threshold (Chen, 2007). Although the jump in SRMR from configural (0.043) to metric (0.086) is larger than the frequently cited  $\Delta SRMR \leq 0.01$  guideline, many researchers place greater emphasis on  $\Delta CFI$  and  $\Delta RMSEA$  (Cheung and Rensvold, 2002; Chen, 2007). Thus, metric invariance is supported.

### Scalar invariance

Adding intercept constraints produced a CFI of 0.934, RMSEA of 0.072, and SRMR of 0.086, with minimal changes from the metric model ( $\Delta CFI = -0.001$ ,  $\Delta RMSEA = -0.002$ ). These values suggest acceptable fit with SRMR just above the recommended cutoff of  $\leq 0.08$ . These results indicate scalar invariance is also

TABLE 2 Multi-group measurement invariance results for rural vs. non-rural STEM teachers.

Step	SB $\chi^2$	df	Scaled $\Delta\chi^2$	$\Delta$ df	$p$	CFI	RMSEA	SRMR	$\Delta$ CFI	$\Delta$ RMSEA	$\Delta$ SRMR
Configural invariance	357.76	194				0.943	0.071	0.043	—	—	—
Metric invariance	389.46	206	32.31	12	0.001	0.936	0.073	0.086	−0.007	+0.002	+0.043
Scalar invariance	406.87	218	15.70	12	0.206	0.934	0.072	0.086	−0.002	−0.001	0.000
Residual invariance	437.63	234	30.56	16	0.015	0.929	0.072	0.081	−0.005	—	−0.005

upheld because the fit indices suggest that fixing intercepts does not significantly degrade model fit.

Residual (strict) invariance

Finally, constraining item-level residuals yielded a CFI of 0.929, RMSEA of 0.072, and SRMR of 0.081. These values suggest acceptable fit with SRMR just above the recommended cutoff of  $\leq 0.08$ . The changes from the scalar model ( $\Delta$ CFI = −0.008,  $\Delta$ RMSEA = +0.002,  $\Delta$ SRMR = −0.005) remained within or near commonly accepted benchmarks. Hence, strict invariance is likewise supported.

Group comparisons of latent means

After establishing scalar invariance, we compared latent means across groups to assess whether rural and non-rural teachers differed in their perceptions of PD resources. Latent means for the non-rural group were constrained to zero, and means for the rural group were freely estimated. Results indicated that rural teachers scored significantly higher on feasibility (estimate = 0.321, SE = 0.11,  $p$  = 0.007) and appropriateness (estimate = 0.288, SE = 0.11,  $p$  = 0.013), but did not differ significantly on local relevance (estimate = 0.096, SE = 0.11,  $p$  = 0.447) or usability (estimate = 0.209, SE = 0.11,  $p$  = 0.083). These results suggest rural teachers perceived PD resources as more feasible and worthwhile relative to their non-rural counterparts, by approximately 0.29–0.32 standard deviations.

Qualitative analysis

Of the 398 survey respondents, 157 provided feedback to the open-ended question, “Is there anything you would like to share about the value or usefulness of the [PD program] resources?” Of these, 59% ( $n$  = 92) identified their school as situated in a rural community and 41% ( $n$  = 65) identified their school as situated in an urban or suburban community.

Depending upon the number of topics addressed, a number of different codes may have been applied to any given response. Table 3 provides the total number of references coded for each of the coding schemes broken out by geographic location of the respondent.

As seen in Table 3, there were small to moderate differences in the content of responses from rural and non-rural teachers. Through the process of open coding, we identified responses related

TABLE 3 Codes applied to responses to the open-ended question for rural vs. non-rural STEM teachers.

Codes	Percent of rural teacher responses	Percent of non-rural teacher responses
Open codes		
Accessibility	5%	9%
Features of the PD	70%	61%
Taking it Back	25%	30%
Deductive codes		
Feasibility	31%	35%
Usability	28%	43%
Appropriateness	33%	21%
Local relevance	5%	11%
Generic response	3%	0%
Sentiment		
Very positive	47%	24%
Moderately positive	23%	17%
Moderately negative	20%	32%
Strongly negative	9%	27%

to three themes: accessibility of PD and PD resources, features of PD and PD resources, and an *in vivo* code of “Taking it Back” to the classroom, the largest difference appearing in the Features of the PD code. Responses under theme of accessibility relate to availability or cost workshops, e.g., “It would be nice to have more options - or changing locations” (rural respondent) and “the discount for teachers should be given to all state certified teachers, not just those currently working for public schools” (non-rural respondent). Responses under the theme of features of PD or PD resources relate to the workshops: activities, content, instructors, resources provided, and the opportunity to experience PD. Examples of these responses are: “It was valuable but geared more toward elementary school and I teach secondary” (non-rural respondent), and “I love the networking!” (rural respondent). Taking it Back responses relate to what happens when the teachers return to their classrooms, e.g., “workshops are awesome and I always say that I’m going to use the materials...but then I run into planning and preparation issues” (rural respondent) and “I would do more if we were able to have more field trips/excursions at our school” (non-rural respondent).

Similarly, we applied social validity codes at somewhat comparable rates for the rural and non-rural teachers’ responses, with the largest differences appearing in the Usability and

Appropriateness codes (rural teachers were more often commenting on appropriateness and less often commenting on usability). Further description of these responses is provided in the integration of qualitative findings and quantitative results.

However, the sentiment of the responses between groups varied considerably. Responses from rural teachers were more often coded as positive than negative (71% positive and 29% negative), while those from non-rural teachers were more often coded as negative than positive (41% positive and 59% negative). Examples of positive responses are: “I have attended 3 years and possibly four (bad memory!) and I still use many of the resources that I learned about and obtained” (rural respondent) and “I value these resources a lot-I get a lot of ideas for my classroom from these resources” (non-rural respondent). Examples of negative responses are: “I just have a really hard time finding stuff that will help in upper division math” (rural respondent) and “resources I received were not tied directly to the standards I teach...the other unfortunate problem was that I didn’t receive enough resources to even have students work in a small group during centers or in a rotation” (non-rural respondent).

In summary, the feedback provided by rural and non-rural teachers in responses to the open-ended question was comparable, though feedback from rural teachers was a little more often about features of the PD or PD resources, little more often about the appropriateness or value given their classroom contexts, and more often positive.

## Integration of qualitative findings and quantitative results

Next, we use examples of teachers’ responses (Table 4) to integrate the findings of this coding with the results of the quantitative analysis. Quotes in Table 4 illustrate the convergence

or divergence with the quantitative results and are not meant to be representative of the entire set of responses to which a particular code was applied.

## Feasibility

While significant mean differences were found for rural and non-rural respondents in regard to feasibility, differences are less apparent in the qualitative analysis. The teachers who took the time to respond to the open-ended question expressed similar concerns about feasibility. The most common topic for both sets of teachers was alignment to their grade level, curriculum, or content standards, and these responses most often co-occurred with the theme Taking it Back and the code “moderately negative.” Rural teachers more often described this as an issue of alignment with grade level, while non-rural teachers more often described this as an issue with curricular or standards alignment. Additionally, there were two distinct, though infrequent, topics. Some rural respondents described how they had successfully applied the resources in informal settings, such as afterschool programs or libraries. Some non-rural respondents discussed access issues, such as cost or offering enough workshops given the demand.

## Appropriateness

The quantitative analysis also revealed significant mean differences for the appropriateness factor, and the qualitative analysis of the open-ended responses does offer convergent evidence for differences in teachers’ assessments of appropriateness. Under the theme of Features, rural teachers more often described the value of the program as a source of professional learning

TABLE 4 Example open-ended responses from rural vs. non-rural STEM teachers.

Deductive code	Rural teacher responses	Non-rural teacher responses
Feasibility	“I have a really hard time finding stuff that will help in upper division math.” “Some material is not for my grade but I’m able to adapt and use resources. Would love to find resources in Spanish.”	“The materials I received are great... I am working on how to best fit them in with my instruction because they do not necessarily go with the curriculum.” “I teach 5 <sup>th</sup> grade and they were not on topic with the standards I teach... I signed up late and that was all that was available at that time for a course to enroll in.”
Usability	“I am working toward creating STEM programs at my school. The [PD program] is providing me with knowledge and equipment that I will utilize.” “There are not always enough kits and such for all enrolled in the classes.”	“Great hands-on and easy to gather resources.” “‘Sample’ kits are essentially useless—without a full class set, they do not get used. Fortunately, this is not usually a problem.”
Appropriateness	“Not only did I get a lot of resources from [PD program] but I was also inspired to implement other components into different areas of my curriculum and units. It ignited by creativity.” “I felt like I was able to develop the skills and knowledge necessary to help my students learn. I can’t say enough about how grateful I am for this program, and the positive impact it has had on the children I serve.”	“I value these resources a lot – I get a lot of ideas for my classroom from these resources.” “I loved going to the [PD program] classes and felt it was beneficial to my students.”
Local relevance	“I feel so very lucky to bring these materials and STEM opportunities back to my community. I would like to connect more opportunities... maybe GPS and different career applications with drones.”	“I find value in exploring how STEM can be found in many different career opportunities, especially ones I had not thought about!”

(15% of the rural response codes as compared to 5% of the non-rural responses). While both groups of respondents commented on the value of the resources (co-occurring with the theme Features) and the enjoyment or engagement by students (co-occurring with the theme Taking it Back), there was a noticeable difference in the sentiment of these responses. Rural teachers were often more effusive and enthusiastic in their comments about the different forms of value the program offers than non-rural teachers.

## Usability

The qualitative findings converge with the results of no significant mean difference between rural and non-rural teachers for the usability factor. For both groups, responses related to usability most often co-occurred with a code under the theme Features. This code addressed the physical resources provided by the workshops, such as activity kits or hands-on materials. Sentiment was similar for these responses. Positive responses related to usability indicate well-designed resources that meet the needs of many teachers for their classroom instruction, and the less positive responses converge on the issue of needing a full class set of resources. Many workshops provided full class sets of materials, and rural and non-rural teachers expressed appreciation for this. There were some workshops that only provided a sample kit, and teachers from both groups communicated that this was not helpful, often citing limited resources to buy their own as the reason.

## Local relevance

The qualitative findings converge with the results of no significant mean difference between rural and non-rural teachers for the local relevance factor. For responses from both groups of teachers, among the deductive codes, this code was applied least often at three references for rural teachers and three references for non-rural teachers. Topics addressed by were the usefulness of specific topics for students' future goals (e.g., cheese making, drones) or careers more generally. These responses did not address the local environment or culture. Sentiment was similar for the responses coded to local relevance.

## Discussion

Drawing on social validity theory (Wolf, 1978) and existing research on PD in diverse settings (Desimone, 2009; Penuel et al., 2007), this study set out to examine whether the four constructs underlying STEM PD resource use (i.e., feasibility, usability, appropriateness, and local relevance) hold comparable meaning for rural and non-rural STEM educators. Establishing such measurement equivalence is critical, as educational practices and resource availability often differ dramatically between rural and non-rural contexts (Howley and Howley, 2005; Harmon, 2020). If an instrument does not measure these constructs in a comparable

manner, differences in scores might erroneously be attributed to contextual factors, rather than reflecting genuine differences in teachers' perceptions (Rutkowski and Svetina, 2014).

To address this, we used a multi-group confirmatory factor analysis, testing configural, metric, scalar, and residual invariance (Cheung and Rensvold, 2002; Chen, 2007). Our results showed that all levels of invariance held. These invariance results suggest that the four constructs operate consistently across groups, ensuring that subsequent comparisons of latent means reflect actual differences rather than psychometric bias. These results parallel previous calls for rigorous multi-group validation when assessing PD resources in varied educational settings (Rutkowski and Svetina, 2014).

Having established scalar invariance, we tested for latent mean differences between rural and non-rural teachers. The analysis revealed that rural educators scored significantly higher on feasibility and appropriateness, whereas no significant differences emerged for local relevance or usability. Effect sizes were moderate (around 0.29–0.32 SD above the non-rural group), suggesting rural teachers perceived implementation to be somewhat easier and more worthwhile than their non-rural counterparts.

In several ways, our qualitative findings converged with these quantitative results. Among those who took the time to respond to the open-ended question, there was a noticeable difference in the overall sentiment of rural teachers and non-rural teachers toward the PD program, with rural teachers' responses coded as moderately or very positive at a higher rate. In responses related to appropriateness, this pattern was particularly prominent. For example, a higher proportion of rural teachers commented very favorably on the opportunity for professional learning that the PD program's workshops offered, providing longer responses and describing how this knowledge continued to be beneficial to their teaching. Also, a nuance in responses related to feasibility may offer some insight into significant mean differences for this factor. Among rural teachers' responses, the issue of alignment to grade level was more frequent, whereas alignment to curriculum or standards was seen more often in non-rural teachers' responses. Perhaps a focus on alignment with curriculum and standards reflects a pressure for fidelity or less autonomy than a focus on meeting the learning needs of students at a particular grade level. If such is the case, teachers who perceive less autonomy may rate the feasibility of a program somewhat lower. Finally, responses related to usability and local relevance were similar regarding topical focus (i.e., most often focused on a class set of materials and content related to local career opportunities), and sentiment was similar as well.

These findings dovetail with prior literature describing the educational experience of rural teachers. In our sample, appropriateness was deemed higher among rural participants, possibly because rural participants found these PD resources beneficial enough to justify time and effort, particularly in a geographic region where content-specific professional development opportunities can be scarce or disconnected from the local and/or school context (Biddle and Azano, 2016; Howley and Howley, 2005; Player, 2015). Similarly, feasibility was higher among rural participants, a finding that is in accord with literature



indicating that small, cohesive rural districts can sometimes afford more flexible, community-driven approaches to integrating new practices (Penuel et al., 2007; Harmon, 2020).

Conversely, usability and local relevance did not differ significantly. Given that usability addresses support for a teacher's classroom implementation, our results converge with the theorization of this as an aspect of social validity that is unique to individuals and distinct from affordances or barriers in the educational or geographic context (Weiner et al., 2017). Given the differing contexts of rural and non-rural schools, one might find a lack of difference in the local relevance of the PD resources surprising. However, this speaks to the importance of PD programs which are tailored to or adaptable across contexts, as was the one which this sample of teachers was evaluating (Biddle and Azano, 2016; Penuel et al., 2007).

## Implications for practice and policy

Social validity is an important consideration in efforts to modify and improve educational programs (Wolf, 1978). These findings offer actionable guidance for both policymakers and PD providers aiming to support STEM teachers in diverse contexts. In making decisions regarding funding or allocation of resources, policymakers can consider how contextual factors such as feasibility and local relevance might be related to differential uptake of programs. Further, when offering PD at scale, policymakers might consider prioritizing programs that offer a structured system for delivering PD in multiple locations with content tailored to grade bands and regional or local contexts, thereby enhancing the perceived value of the programs across diverse teaching contexts.

Because rural educators in this study reported higher feasibility and appropriateness of PD resources and comparable local relevance, developers can leverage that strength by embedding place-based examples or local community ties, which align well with small-district collaboration networks (Harmon, 2020). This might mean designing resource modules that explicitly reference local industries or community events, allowing rural teachers to demonstrate clear relevance to their student population. At the same time, non-rural educators, who may navigate more complex district policies or larger student populations, could benefit from PD guidance on streamlining the adoption of new resources and integrating them into a system's existing priorities.

In both cases, tangible recommendations involve fostering strong local partnerships and providing customizable PD elements that educators can adapt to their unique school environments. For educational leaders, the results underscore that "one-size-fits-all" PD approaches may overlook critical contextual nuances. District or state officials might consider pilot-testing materials in both rural and non-rural settings to ensure that language, examples, and pacing resonate with each population (Desimone, 2009). Additionally, reinforcing teacher communities of practice, particularly across regional boundaries, could facilitate resource sharing and collaboration, thereby boosting feasibility and utility for all.

## Limitations and future directions

This study faced several limitations. First, there is potential self-selection bias, as teachers who opted into this STEM PD (and later responded to our survey) may be more motivated or better-resourced than non-responders. Second, generalizability beyond this particular state PD program or population from which this sample is drawn remains an open question, especially given the unique demographics and policy context that may shape teacher experiences (Howley and Howley, 2005). Third, reliance on self-report data can introduce biases tied to social desirability or recall. Lastly, while the qualitative responses helped contextualize quantitative patterns, they were limited in length and depth, meaning that certain nuances might be underexplored.

Building on the study's results, future research could explore measurement invariance of STEM PD resource surveys across additional states or subject areas, such as mathematics vs. science-focused PD, to see if the same constructs and group differences hold. It would also be valuable to conduct longitudinal research that tracks how teachers' perceptions of feasibility, usability, appropriateness, and local relevance evolve over time or after multiple PD experiences (Desimone, 2009).

Moreover, adopting more extensive qualitative methods, including focus groups or one-on-one interviews, would enable a deeper dive into place-based factors influencing PD uptake. Such methods could help identify more complex or unexpected barriers and facilitators that do not emerge in brief written responses. Finally, future research might also incorporate classroom observations or student outcome data to connect teacher perceptions directly with evidence of instructional changes or student learning gains. By embracing these lines of inquiry, the field can continue refining PD resources and strategies that are responsive to the distinct realities faced by both rural and non-rural educators.

## Conclusion

This study is an important step toward understanding the social validity of STEM professional development resources in both rural and non-rural educational contexts. By employing a mixed-methods approach, we aim to bridge the gap between the unique challenges faced by rural educators and the universal principles of effective PD. The MG-CFA results provide preliminary evidence that the STEM PD Resource Use Survey measures similar constructs across rural and non-rural settings, which is crucial for ensuring that comparisons between these groups are meaningful and valid.

The findings of this study have the potential to inform the design and implementation of STEM PD initiatives that are more sensitive to the unique needs of educators in diverse contexts. By understanding both the universality and the context-specific factors that influence teachers' perceptions of PD resources, we can better support educators in effectively implementing STEM practices in their classrooms. Ultimately, this research contributes to the broader goal of ensuring that all educators, regardless of their geographic location, have access to high-quality, relevant, and impactful professional development opportunities.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Office of Research Compliance, Human Subjects Board, Boise State University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

AC: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – original draft, Writing – review & editing. AS: Formal analysis, Funding acquisition, Methodology, Resources, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was

supported in part by the Institute for Measurement Methodology in Rural STEM Education (IMMERSE-NSF#2126060) through NSF EHR Core Research: Building Capacity in STEM Education Research.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. Generative AI was used to produce an early draft of the abstract.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (2014). Standards for educational and psychological testing. American Educational Research Association.
- Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*. New York, NY: Guilford Publications.
- Biddle, C., and Azano, A. P. (2016). Constructing and reconstructing the “rural school problem”: a century of rural education research. *Rev. Res. Educ.* 40, 298–325. doi: 10.3102/0091732X16667700
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Publications.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equation Model. Multidiscip. J.* 14, 464–504. doi: 10.1080/10705510701301834
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equation Model.* 9, 233–255. doi: 10.1207/S15328007SEM0902\_5
- Creswell, J. W., and Plano Clark, V. L. (2017). *Designing and Conducting Mixed Methods Research, 3rd Edn.* Thousand Oaks, CA: Sage Publications.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: toward better conceptualizations and measures. *Educ. Res.* 38, 181–199. doi: 10.3102/0013189X08331140
- Greer, F. W., Wilson, B. S., DiStefano, C., and Liu, J. (2012). Considering social validity in the context of emotional and behavioral screening. *Sch. Psychol. Forum Res. Pract.* 6, 148–159.
- Harmon, H. L. (2020). Innovating a promising practice in high poverty rural school districts. *Rural Educ.* 41, 26–42. doi: 10.35608/ruraled.v41i3.1018
- Holmbeck, G. N., and Devine, K. A. (2009). Editorial: an author's checklist for measure development and validation manuscripts. *J. Pediatr. Psychol.* 34, 691–696. doi: 10.1093/jpepsy/jsp046
- Howard, M., Alexiades, A., Schuster, C., and Raya, R. (2024). Indigenous student perceptions on cultural relevance, career development, and relationships in a culturally relevant undergraduate STEM program. *Int. J. Sci. Math. Educ.* 22, 1–23. doi: 10.1007/s10763-023-10360-3
- Howley, A., and Howley, C. (2005). High-quality teaching: providing for rural teachers' professional development. *Rural Educ.* 26, 1–5. doi: 10.35608/ruraled.v26i2.509
- Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Publications.
- Lakin, J. M., and Shannon, D. M. (2015). The role of treatment acceptability, effectiveness, and understanding in treatment fidelity: predicting implementation variation in a middle school science program. *Stud. Educ. Eval.* 47, 28–37. doi: 10.1016/j.stueduc.2015.06.002
- O'Neill, T., Finau-Faumuina, B. M., and Ford, T. U. L. (2023). Toward decolonizing STEM: centering place and sense of place for community-based problem-solving. *J. Res. Sci. Teach.* 60, 1755–1785. doi: 10.1002/tea.21858
- Penuel, W. R., Fishman, B. J., Yamaguchi, R., and Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *Am. Educ. Res. J.* 44, 921–958. doi: 10.3102/0002831207308221
- Player, D. (2015). *The Supply and Demand for Rural Teachers*. Rural Opportunities Consortium of Idaho.
- Rossee, Y., Jorgensen, T. D., and Rockwood, N. (2023). *lavaan: latent variable analysis... R package version 0.6-16*. Available online at: <https://lavaan.ugent.be> (Accessed October 12, 2024).
- Rutkowski, L., and Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educ. Psychol. Meas.* 74, 31–57. doi: 10.1177/0013164413498257
- Stuckey, M., Hofstein, A., Mamlok-Naaman, R., and Eilks, I. (2013). The meaning of ‘relevance’ in science education and its implications for the science curriculum. *Stud. Sci. Educ.* 49, 1–34. doi: 10.1080/03057267.2013.802463

- Villarreal-Zegarra, D., Copez-Lonzoy, A., Bernabé-Ortiz, A., Melendez-Torres, G. J., and Bazo-Alvarez, J. C. (2019). Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): a measurement invariance study across groups by demographic characteristics. *PLoS ONE* 14:e0221717. doi: 10.1371/journal.pone.0221717
- Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., et al. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Sci.* 12, 1–12. doi: 10.1186/s13012-017-0635-3
- Wolf, M. M. (1978). Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. *J. Appl. Behav. Anal.* 11, 203–214. doi: 10.1901/jaba.1978.11-203
- Xu, L., Fang, S. C., and Hobbs, L. (2023). The relevance of STEM: a case study of an Australian secondary school as an arena of STEM curriculum innovation and enactment. *Int. J. Sci. Math. Educ.* 21, 667–689. doi: 10.1007/s10763-022-10267-5
- Yoon, M., and Lai, M. H. (2018). Testing factorial invariance with unbalanced samples. *Struct. Equation Model. Multidiscip. J.* 25, 201–213. doi: 10.1080/10705511.2017.1387859