



Validation of the teaching equity enactment scenario scale in Singapore: a mixed-methods convergent study

Wen-Chia Chang¹ 

Accepted: 2 November 2022 / Published online: 7 January 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Given the persistent and endemic educational inequalities and increasingly diverse and yet politically divisive societies, teaching that is inclusive for all students with a commitment to recognizing and seeking ways to challenge systemic inequity is one approach to addressing persistent disparities. To support teachers' professional growth and provide evidence for research/evaluation on teacher learning, this study validates the Teaching Equity Enactment Scenario (TEES) Scale, an existing instrument first conceptualized and developed in mostly Western contexts, among Singapore teachers. A mixed-method design, integrating a survey of 78 teachers and a follow-up interview using the think-aloud technique to illicit pedagogical practices of five survey participants, is used. The quantitative results indicate that the TEES Scale measures a unidimensional construct of enacting equity-centered teaching practice from the lower to the higher level as hypothesized and can provide reliable and meaningful interpretations of participants' scores. The qualitative results provide contextualized information about participants' survey experiences and the patterns of practice among higher- and lower-scoring teachers. Specifically, despite the common boundaries, structure, and parameters that condition Singapore teachers' work, higher- and lower-scoring teachers diverge in their views of learners, knowledge and knowledge construction, perceived professional roles and identities, and instructional practices in important manners. Findings of the two components confirm each other and offer a fuller picture of the degree to which the TEES Scale can provide reliable and meaningful information about Singapore teachers' practice for equity for the intended uses. Limitations and future studies are discussed.

Keywords Validation · Mixed methods · Equity · Teaching · Teacher learning · Rasch measurement

✉ Wen-Chia Chang
wenchia.chang@nie.edu.sg

¹ Nanyang Technological University, National Institute of Education, NIE2-03-59, 1 Nanyang Walk, Singapore 637616, Singapore

1 Introduction and background

Educational inequality in achievement, resources, and opportunity to learn based on class, race/ethnicity, gender, linguistic background, and other factors is a persistent and complex international issue (UNESCO 2015). Singapore is no exception, despite the nation's top performances on large-scale international assessments (Tan 2007, 2014). Across the globe, teacher quality has also become a policy focus and a lever for addressing educational challenges (Akiba and LeTendre 2018; Goodwin and Low 2021). In Singapore, scholars have increasingly argued that teaching that recognizes and affirms students' culture and seeks ways to resist dominant norms about and rules for so-called low-achieving students can offer possibility of addressing educational disparities (e.g., Alviar-Martin and Ho 2011; Heng and Atencio 2017; Lim and Tan 2018).

With increasing attention to educational (in)equity in Singapore (Kwek et al. 2019; Tan 2019), an instrument that can provide reliable and meaningful information about teachers' equity-centered practices can be used to support teachers' professional growth and to allow for further inquiries into how, to what extent, and under what conditions teachers learn to enact equity-centered teaching practices. A recent review of existing, relevant assessment tools suggested that most of the assessment tools were intended to capture teachers' mental schemata, such as teachers' beliefs and attitudes, rather than practices and that very few instruments are designed to measure constructs of teaching practices for equity and/or social justice commitment (Chang and Cochran-Smith 2022).

Considering the pressing need in the field, the Teaching Equity Enactment Scenario (TEES) Scale was first developed to capture the complexity of enactment of equity-centered teaching practice among educators in the U.S. and New Zealand (Chang et al. 2019), contexts wherein the historical foundations, political contexts and structures, and social organizations are different from those in Singapore. The results of the initial TEES Scale development provided robust evidence to support the reliability and validity of the scale (Chang et al. 2019). However, given the context in which the scale was developed, a validation study seeking to investigate whether and how the TEES Scale scores can be interpreted and used appropriately in the Singapore context is needed. In the following sections, I first present the TEES Scale, including the construct theories and measurement approach that guided the scale development. I then briefly discuss Singapore's educational context, which informs the need for a validation study using a mixed-method research approach.

1.1 The TEES scale

The construct of practice for equity was grounded in the theories of equity-centered teaching (Cochran-Smith et al. 2016) and informed by purposefully selected five international syntheses¹ about teaching practices that contribute to broadly defined student learning, have a positive impact on learners of diverse cultural and linguistic backgrounds, and reflect a complex view of teaching. Through a content analysis of the literature, six principles of practice for equity that define the TEES construct were identified: (1) selecting

¹ The international literature includes three New Zealand Best Evidence Syntheses (Aitken and Sinnema 2008; Alton-Lee 2003; Anthony and Walsaw 2007), the Teaching and Learning Research Project in the U.K. (James and Pollard 2006), the Measures of Effective Teaching in the U.S. (MET Project 2013), Te Kotahitanga Effective Teaching Profile (Bishop et al. 2009), and the Center for Research on Education, Diversity, and Excellence's five standards for effective pedagogy (Dalton 2007).

worthwhile content and designing and implementing learning opportunities aligned with valued outcomes; (2) connecting to students as learners and their lives and experiences; (3) creating learning-focused, respectful, and supportive learning environments; (4) using evidence to scaffold learning and improve teaching; (5) taking an inquiry stance for further professional engagement and learning; and (6) recognizing and challenging classroom, school, and societal practices that reproduce inequity (Chang et al. 2019). Relevant empirical evidence has suggested that the six principles are interconnected (Grudnoff et al. 2017), meaning that it is the enactment of multiple principles as a whole, rather than the enactment of a single principle, that enhances student learning.

The TEES Scale was developed by applying the Rasch/Guttman-based Scenario (RGS) scale approach (Ludlow et al. 2020, 2021; Ludlow et al. 2014), integrating Rasch measurement (Rasch 1960/80) and Guttman facet theory (Guttman 1959) to construct scenario-style items that measure a unidimensional construct. Rasch measurement principles -- unidimensionality, variability, hierarchical continuum, even spread, equal discrimination, local independence, and theory/data confirmation (Wright and Masters 1982) -- guide the scale development. Guttman's facet theory design and mapping sentence technique (Borg and Shye 1995; Guttman and Greenbaum 1998) facilitate the systematic construction of the scenarios. I briefly describe the process of developing the TEES Scale below and refer readers to Chang et al. (2019) and Chang (2021) for the full, detailed development procedure.

Guided by the RGS approach, an iterative content analysis of the selected literature led to the identification of the six interconnected principles (or "facets," a term used in Guttman facet theory to indicate distinct and yet interrelated characteristics defining a construct) of practice for equity that together define the enactment of equity-centered practice as a single, unidimensional construct. Further, guided by the Rasch principles, the enactment of equity-centered practice was hypothesized as a ladder-like construct ranging from lower to higher levels of enacting equity-centered practice. The lower level of practice for equity portrays teaching from a technical view -- transmitting standardized, seemingly value-neutral knowledge to students regardless of their backgrounds, seeing students through a deficit lens, and assuming no responsibility to advocate on behalf of their students and challenge inequitable norms and practices. Moving along the construct continuum, the higher level of practice for equity reflects a more complex view of teaching: having a critical view of knowledge, seeing students through an asset-based lens, co-constructing knowledge with students through connecting to their life and prior experiences, and developing critical consciousness and seeking ways to challenge inequities. In the initial construct mapping stage, rich, narrative description for each facet was first developed. With an in-depth understanding of the facets, narrative descriptions capturing hierarchical variations representing low, moderate, and high levels within each facet were then generated.

Next, given the six facets and three levels within each facet, rather than using all possible combinations ($3^6 = 729$), two decisions were made to select and assemble facet level descriptions to form scenarios. First, I only selected facet descriptions from just the highest, the moderate, and the lowest to construct plausible scenarios that capture three distinct levels along the construct continuum. This decision was based on the understanding of the construct and a need to obtain a proof of concept first before trying to capture the subtleties in between. Second, it is impossible to include all six facets in one scenario because participants would be overwhelmed. I decided that three facets would be feasible and that Facet 6, *recognizing and challenging inequities*, must be present in all scenario combinations, aligning with the construct theory. As such, Facets 1 to 5 are to be systematically selected so that some of the five facets overlap between scenarios. As shown in Table 1,

Table 1 Structure for scenario development

Construct theory	Level of enactment	Scenarios	Facets of Teaching for equity								
			1	2	3	4	5	6			
Higher-level enactment: Complex view of teaching Asset-based view of learners’ lived experience and culture A constructivist view of knowledge and knowledge construction Teaching is value-laden and inherently political; teachers are agents of change	High	F126 HHH	X	X					X		
		F236 HHH		X	X				X		
		F346 HHH			X	X			X		
		F456 HHH					X	X	X		
		F156 HHH	X					X	X		
	Moderate	F126 MMM	X	X					X		
		F236 MMM		X	X				X		
		F346 MMM			X	X			X		
		F456 MMM					X	X	X		
		F156 MMM	X					X	X		
		Lower-level enactment: Technical view of teaching Deficit lens of learners’ lived experience and culture A static view of knowledge and knowledge construction Teaching is value-free, and a teacher’s responsibility excludes advocacy	Low	F126 LLL	X	X					X
				F236 LLL		X	X				X
				F346 LLL			X	X			X
				F456 LLL					X	X	X
				F156 LLL	X					X	X

X indicates that the specific facet is included in the scenario

this approach resulted in five scenarios capturing each of the three levels along the construct continuum. For instance, Scenarios F126HHH (i.e., Facets 1, 2, 6, “H”ighest level), F236HHH, F346HHH, F456HHH, and F156HHH capture the high level of enactment of practice for equity. Scenarios based on the same facet combinations were developed to capture the Moderate (M) and the Lower (L) levels of enactment of practice for equity. Altogether, these decisions resulted in a total of 15 scenarios to be developed next.

Once the scale structure is determined, the Guttman’s mapping sentence technique (Borg and Shye 1995; Guttman and Greenbaum 1998) facilitates the construction of scenarios. A mapping sentence is a grammatical device that consists of the formal elements – descriptions that capture the levels of selected facets – and informal elements – phrases that remain constant to link descriptions of facets and provide a meaningful context. Each scenario encompasses four to five sentences and a teacher’s name, which together portray the hypothetical teacher’s classroom practice. To respond to each scenario as an item,

participants are instructed to reflect on their practice, compare it to the practice described in a scenario, and indicate their own practice based on a five-point scale (*Much lower*=1, *Slightly lower*=2, *About the same*=3, *Slightly higher*=4, *Much higher*=5). For instance, choosing *About the same* indicates that respondents align their practices with the practice captured in a scenario. Informed by the construct theories, it is expected that scenarios capturing the higher level of enactment of practice for equity would be the most difficult for participants to choose. About the same or the *higher* response options, and scenarios capturing the lower level of enactment of practice for equity would be the easiest for participants to choose the *higher* response options. Higher scores indicate higher levels of enactment of equity-centered practice. Empirical evidence based on p*About the same* participants in the U.S. and New Zealand confirmed the Rasch model expectations, providing support for the reliability and validity of the TEES Scale (Chang et al. 2019; Chang 2021).

1.2 Teaching for equity in the Singapore context

Singapore's unique state of education is closely connected to the country's historical, political, and economic contexts. As a city state with limited natural resources and a diverse population comprised of multiethnic, multilingual, and multireligious groups, maintaining social cohesion and preparing highly skilled human capital to ensure national economic prosperity and security in global competition have been the state's priorities. To this end, ideas of multiculturalism and meritocracy are two pillar principles promulgated by the state in informing policies and practices, including education, which plays an instrumental role in nation building (Gopinathan 2015).

Given Singapore's diverse population, the idea of multiculturalism – equal treatment to all regardless of race, gender, religion, or linguistic background – has been advocated by the state and its founding leaders to promote tolerance, build social cohesion, and strengthen loyalty among citizens. The National Education curriculum launched in 1997 is one of many efforts to instill national identity by teaching values of multiculturalism and meritocracy. Within the purview of multiculturalism, although issues related to race, religion, and language are covered and discussed in the curriculum, the approach to these issues is largely superficial –i.e., learning about the different customs and traditions of diverse communities and how diversity contributes to Singapore's nation building, rather than delving into the real and important differences in values, beliefs, and world views (Bokhorst-Heng 2007). Studies have found that the local laws and official discourse on multiculturalism confine teachers' perceptions regarding what classroom discourses are deemed acceptable or not (Alviar-Martin and Ho 2011; Heng and Lim 2021; Ho et al. 2014). Critiques have also argued that the official discourse on multiculturalism is a "repressive device" by the government to maintain social control (Bokhorst-Heng 2007; Chua 2003).

Moreover, the idea of meritocracy—individuals' hard work and/or talent as the sole determinants of one's success, assuming a level playing field and standardized tests as objective and value-neutral yardsticks to determine one's merits—was propagated to promise nondiscrimination and equal treatment for all regardless of background (Lim 2013; Ho 2021; Tan 2008, 2014). In education, the academic tracking practice (or "streaming" in Singapore), according to which students are sorted into classes starting in 4th grade and again as students leave primary for secondary schools based on their standardized test scores, is a prominent example of how meritocracy shapes education practices. With the persistent educational disparities along race/ethnicity and class lines, the idea of meritocracy is also used by the state to legitimize the highly stratified social and educational opportunities and

outcomes (Lim 2013; Tan 2008, 2019; Talib and Fitzgerald 2015). Critiques have pointed out that the meritocratic system not only fails to recognize pre-existing structural, unequal conditions, but it also perpetuates inequality (Abu Bakar 2009; Tan 2008, 2019). Moreover, studies have shown how the discourse of meritocracy shapes teachers' perceptions of learners, learning, and teaching (e.g., Anderson 2015; Heng and Atencio 2017; Heng and Lim 2021).

In Singapore, teachers are prepared and expected to play a key role in nation building (Copinathan 2015). The unique educational context largely frames the discourses on what is considered high-quality, effective teaching. In examining the policy discourse on teacher professionalism in Singapore, Ro (2020) concluded that Singapore's policy discourses on teacher professionalism mostly emphasize teachers' roles in delivering and adapting a common, practical knowledge base for teaching so that teachers become effective in implementing the national curriculum and fulfilling the national agenda. Missing are discourses that consider teachers to be agents of change for social good, adopt a critical perspective on curricula, and exercise professional judgment in their work. Recent studies have also suggested that teachers' ideas about and practices pertaining to diversity and equity are socially constructed and shaped by Singapore's unique sociopolitical forces and must be understood in context (e.g., Alviar-Martin and Ho 2011; Heng and Lim 2021). Altogether, Singapore's unique education system, which frames how teachers perceive their professional identities, roles and responsibilities, and practices, poses questions regarding whether and to what extent the TEES Scale can reliably and meaningfully differentiate Singapore teachers' levels of enacting equity-centered practice.

1.3 Current study

This study aims to understand whether and how the scores derived from the TEES Scale accurately represent participants' differential endorsement of the scenarios and can be interpreted appropriately to infer Singapore teachers' varying levels of enacting practice for equity in the classroom for purposes of self-reflection and professional learning and to provide evidence for research/evaluation studies. Recognizing the complexity of the validation task and the need to include the lived experience of Singapore teachers, a mixed-method research approach encompassing a quantitative survey component and a follow-up interview with a subset of survey participants was chosen to address the study's purpose. In other words, multiple types of evidence are to be collected, triangulated, and judged to investigate the plausibility of the aforementioned validity claim, a process similar to the logic and work of program evaluation (Kane 2006). The research questions are as follows.

- Quantitative: Do the empirical results confirm the hypothesized unidimensional Rasch measurement model informed by the construct theories?
- Qualitative: How do Singapore teachers describe their survey-taking experience and their classroom practices as they reflect on and respond to the TEES Scale?
- Mixed methods: To what extent do the psychometric results of the scale and interview with the respondents converge or diverge?

2 Methodology

Mixed methods research integrates quantitative and qualitative research approaches into a specific and logical research design for the purposes of “breadth and depth of understanding and corroboration” (Johnson et al. 2007, p. 123). Considering the context, research problem, and purpose of this study, a mixed method is particularly appropriate and useful. Specifically, while quantitative methods and statistical procedures are dominant in the field of measurement, scholars have cautioned that sole reliance on traditional methods and statistical criteria produces necessary, but insufficient, and inadequate validity evidence (Maul 2017). To understand the extent to which the scale scores are meaningful and useful in the Singapore context, a qualitative approach can provide contextualized understanding of the phenomenon, strengthen the information drawn from quantitative methods, and provide a fuller picture of the problem at hand. To this end, pragmatism, a pluralist, inclusive, and complementary stance (Biesta 2010; Morgan 2007), guides this study. From this perspective, research inquiries are ways to develop warranted assertions about the phenomenon of interest and are to be evaluated pragmatically given the research needs, questions, and contexts.

2.1 Research design

This validation study utilizes the fully integrated mixed-methods convergent design, in which both the quantitative and qualitative strands interact during the implementation and occur roughly at the same time (Creswell and Plano Clark 2018). The quantitative strand uses a survey comprising the TEES Scale and relevant background questions. The qualitative strand uses individual interviews with a small subset of voluntary survey participants. The qualitative interview utilizes the technique of a think-aloud exercise, in which participants were asked to describe how and why they selected a certain response option to several scenarios by elaborating on their pedagogical practices and decisions in their contexts. This design is appropriate since the intent is to compare the psychometric results of the scale with the interview texts from a subset of the survey participants to evaluate the validity claims of the TEES Scale. The qualitative interview data triangulate and illustrate the psychometric information of the scale, and together, both strands provide a more complete and complex understanding of the research questions.

2.2 Data collection

The quantitative strand uses a Qualtrics survey that includes the TEES Scale and five questions on participants’ gender, race/ethnicity, years, levels, and subjects taught. Prior to the survey administration, I engaged two content experts to review the survey instructions and the 15 scenarios. The expert review focused on the clarity and comprehensibility of the instructions and the relevance of terms used in the scenarios considering the local context. Revisions were made while maintaining the original meaning/content and the difficulty levels of the scenarios.

The qualitative strand included a one-hour semi-structured interview with each interview participant. The interview questions consisted of two main sections. The first section aimed to understand participants’ understanding and application of the survey instructions

and their reactions to the scenarios (i.e., readability, clarity, respondent fatigue). The second section used the think-aloud technique to elicit participants' self-described pedagogical practices and decisions that they reflected on as they responded to each scenario. Each participant discussed approximately three to five scenarios during the interviews. The participants were encouraged to provide specific examples and elaborate on aspects such as their views of themselves as teachers, learners, teaching and learning, knowledge, instructional designs and choices, and advocacy. In addition, documents such as syllabus and instructional materials were collected to support other data sources.

2.3 Participants

The target participants were teachers who teach any of the academic subjects (i.e., mathematics, science, languages, and humanities) in primary or secondary schools in Singapore. Participants who had some experience teaching academic tracks for students with lower exam scores, i.e., foundation courses in primary schools or Normal Academy and Normal Technical (NA/NT) streams in secondary schools, were encouraged to participate. Participants were recruited through school leaders and instructors in master's-level program courses for in-service teachers at the only teacher preparation institution in Singapore. A total of 85 participants consented and responded to the anonymous survey, among whom five survey participants also volunteered to participate in the interviews.

Before proceeding to the analysis, data quality was checked using participants' median response duration (Buchanan and Scofield 2018; Zhu and Carterette 2010). Seven "speeders" who took less than half of the median duration (i.e., 671 s) and provided repetitive response patterns (e.g., 3s, 4s) were identified and removed. Table 2 presents the participants' profiles. Among the remaining 78 participants, the majority were female (60.3%) and of Chinese ethnicity (60.3%). Most participants teach at the secondary level (60.3%) and have taught for more than 10 years (59.0%). Approximately half of the participants taught languages and humanities subjects, while one third taught science or mathematics.

2.4 Data analysis

Consistent with the convergent design, quantitative and qualitative data are first analyzed independently. A joint display table is then used to bring the two analysis results together for comparison and interpretation.

2.4.1 Quantitative analysis

The Rasch rating scale model (Andrich 1978) is employed to conduct the quantitative analysis, using the WINSTEPS software package (Linacre 2021, v5.1.7). In the rating scale model, the probability of a participant selecting a response option (or category) to a given item is a function of participants' levels of enacting practice for equity (i.e., person ability) and the item's difficulty. Moreover, the increasing difficulty of moving from one response category to the next or to a higher one (e.g., from choosing *About the same* to *Slightly higher*) is presumed to be the same for all 15 items. Participants' ability and items' difficulty are reported in *logit* (log odds) units as a result of the model transforming the raw scores. Positive and higher logits indicate high-scoring individuals or more difficult items, while negative logits indicate low-scoring individuals or easier items (Ludlow and Haley

Table 2 Participant background

Variables	Survey (<i>N</i> = 78) Count (Percentage)	Interview (<i>n</i> = 5)
Gender (<i>n</i> = 72)		
Female	47 (60.3%)	3
Male	19 (24.4%)	2
Prefer not to disclose	6 (7.7%)	0
Race/Ethnicity (<i>n</i> = 72)		
Malay	10 (12.8%)	1
Indian	8 (10.3%)	0
Chinese	47 (60.3%)	3
Eurasian	0 (0.0%)	0
Other	1 (1.3%)	1
Prefer not to disclose	6 (7.7%)	0
Teaching level (<i>n</i> = 71)		
Primary	24 (30.8%)	1
Secondary	47 (60.3%)	4
Years Taught (<i>n</i> = 72)		
3 years or less	5 (6.4%)	0
> 3 years and ≤ 5 years	2 (2.6%)	0
> 5 years and ≤ 10 years	19 (24.4%)	2
More than 10 years	46 (59.0%)	3
Teaching Subject (<i>n</i> = 72)		
Languages	27 (34.6%)	1
Humanities	13 (16.7%)	3
Mathematics	11 (14.1%)	1
Science	12 (15.4%)	0
Other	9 (11.5%)	0

1995). While the sample size is small, Rasch literature suggests that such a sample size can produce item and person measures stable within ± 0.5 logit with 95% confidence (Linacre 1994).

Multiple psychometric indicators are used to evaluate whether the empirical data confirm the Rasch model expectations, providing evidence to support the reliability and validity of the TEES Scale. The psychometric indicators include the following:

- (a) A variable (Wright) map: a variable map provides a visual representation regarding whether the calibrated items are evenly spread to define the difficulty progression of a single variable as hypothesized, providing evidence relating to the scale's internal structure.
- (b) Rasch reliability and separation indices: Rasch person reliability and separation indicate the extent to which participants can be reliably differentiated into different performance levels, while Rasch item reliability and separation indicate the extent to which items

are sufficiently spread out along the construct continuum. Rasch person reliability of at least .8, a person separation index of at least 2.0, and item separation of at least 3.0 are acceptable (Linacre 2016; Wright and Masters 1982).

- (c) Categorical characteristic curves (CCCs): response categories are used by participants as intended and whether the probability of moving from one response category to the next is in an increasingly difficult pattern. Several key criteria are used to check if response categories function as intended: (a) a unimodal shape for each rating scale distribution to show participants' appropriate use of responses, (b) an increasing and ordered pattern of the rating scale category (Andrich) thresholds to show that each category probability curve occupies a unique and wide range, and (c) an increasing pattern of the average person measures associated with each successive response category to show the proper use of rating scale across items (Wolfe and Smith 2007).
- (d) Fit statistics: Fit statistics reflect the discrepancy between the expected and the observed responses. Weighted and unweighted (Infit- and Outfit-MNSQ in WINSTEPS) mean squares of standardized residuals between the observed and the expected responses and the associated t statistics (ZSTD in WINSTEPS) are examined to check how well the calibrated items fit the Rasch model expectations. Infit- and Outfit-MNSQs ranging between 0.5 and 1.5 are considered productive for measurement (Linacre 2002). Large fit statistics often indicate a need for further investigation into participants' response patterns and a need for item revision/deletion.
- (e) Dimension analysis: Rasch unidimensionality assumes that, after conditioning out variance explained by the Rasch model, item standardized residuals do not correlate, and the remaining error variance does not measure a substantially meaningful and significant second dimension that warrants the creation of another test. This assumption is checked by conducting principal component analysis (PCA) on the item standardized residuals. An eigenvalue less than 3.0 is considered acceptable (Wolfe and Smith 2007).

2.4.2 Qualitative analysis

Participant interviews were transcribed immediately afterward and were checked for accuracy with the participants. Data were imported and analyzed using the MAXQDA software (VERBI Software 2019). An inductive coding approach, which allowed codes to emerge through the lived experiences of Singapore teachers, was used in the initial phase (Miles et al. 2013). Multiple types of codes – descriptive, “in vivo”, process, and value coding – were used to sort through data regarding contextual factors (e.g., Singapore's education system), teachers' pedagogical practices, and teachers' beliefs and views about policies, learning and learners, and teaching. Subsequently, the conceptual framework of the equity-centered teaching practices was used to revise and structure the initially emerged codes and to guide the development of major themes and categories during the

second cycle of coding. The initial and second coding phases facilitate the identification of patterns of practices among participants with lower to higher scale scores.

2.4.3 Merged analysis

A joint display table is constructed to bring together the Rasch analysis results of the TEES Scale and the analysis of interview data. The joint display table compares both results side by side and shows the ways in which they confirm, contradict, and/or expand each other. The comparison allows for further consideration of how the quantitative and qualitative results provide insights into the proposed interpretations and uses of the TEES Scale in the Singapore context.

2.5 Validity strategies

Several strategies are adopted to enhance the inference quality of this study. For the quantitative component, the *Standards for Educational and Psychological Testing* (American Educational Research Association et al. 2014) and a validation framework for Rasch models (Wolfe and Smith Jr. 2007) guided the practice. First, a clear specification of construct theory, connection between construct theory and item development, and local expert reviews provided validity evidence based on test content; second, psychometric properties on rating scale functioning and participants' response patterns and fit statistics provided validity evidence on response processes; third, variable maps showing the data/theory match, item fit statistics, and dimensionality analysis provided validity evidence based on internal structure; and, fourth, reliability estimates provided some evidence on generalizability. Guba's (1991) criteria for assessing the trustworthiness of naturalistic inquiries guided the strategies for qualitative component. Specifically, strategies including (a) member checks, (b) triangulation of different data sources, (c) documentation of the data collection, analysis, and interpretation process (i.e., an audit trail), and (d) practicing research reflexivity were used to improve the trustworthiness of interpretations. For the mixed-methods component, guided by strategies for the convergent design (Creswell and Plano Clark 2018), using parallel concepts in data collection and a joint display technique to compare quantitative and qualitative results were approaches adopted to draw correct inferences. While the quantitative and qualitative strands have unequal sample sizes that can pose validity threats to the mixed-methods inference, this study intends to compare groups with a specific range of scale scores with individual experiences elaborated through interviews.

2.6 Positionality

I am an Asian, female, and bilingual scholar from a middle-class family. My experience working with marginalized communities in Asia along with my racialized and gendered educational and professional experiences in the U.S. higher education institutions shape my scholarship. As a critical scholar, my ongoing work to understand how teaching and learning to teach with a commitment to equity and/or social justice can be possible and enacted in Asian contexts informs this study.

3 Findings

This section first discusses the psychometric results of the TEES Scale, followed by the results of the qualitative interviews. A joint display table is then presented to compare and interpret the quantitative and qualitative results.

3.1 Psychometric properties of the TEES scale

3.1.1 Variable map

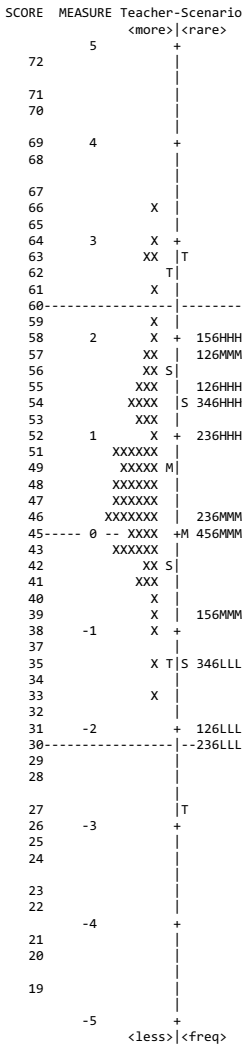
Figure 1 is a variable map showing the locations of individual participants on the left and calibrated items on the right along the central vertical line representing the hierarchical construct continuum from the lower to higher level of enacting equity-centered teaching practice. The “M” on each side indicates the average of person or item measures, which are in both logit units and raw scores. The horizontal lines indicate the average raw scores of 2 (*Slightly lower*), 3 (*About the same*), and 4 (*Slightly higher*). Overall, the empirical data confirm the Rasch model expectation. The difficulty order of the scenarios is mostly as intended. That is, scenarios capturing the higher-level enactment of practice for equity were the most difficult for participants to give “higher” ratings to than the practices captured in the scenarios, followed by the scenarios capturing the moderate level of enactment. Scenarios capturing the lower-level enactment of practice for equity were the easiest for participants to select “higher” responses. The only exception was Scenario F126MMM. Based on the data from the sampled Singaporean teachers, this moderate-level scenario is at a similar difficulty level as other scenarios capturing the higher levels of practice for equity, which was unexpected given the a priori theory.

Using this variable map, participants’ scores can be interpreted with regard to their levels of enactment of practice for equity in classroom, and scenarios provide rich, qualitative interpretations of participants’ scores in relation to the construct. For instance, participants scoring between 52 and 60, they are likely to select *About the same* to the higher-level scenarios next to their locations on the map such as Scenario F346HHH, Siti’s practice, and choose *Slightly higher* to scenarios below their locations. These participants tend to align their practice to a more complex view of teaching: teaching requires recognizing and including assets students bring with them to the classroom, sharing the power of constructing knowledge with students, and using various assessment approaches to scaffold learning and improve teaching. While for participants scoring between 30 and 43, they are likely to select *Slightly lower* to the lower-level scenarios next to their locations such as Scenario F456LLL, Daniel’s practice. Lower-scoring participants tend to align their practice to a more technical view of teaching: teaching is merely transmitting knowledge to students without recognizing and making learning relevant to their backgrounds, lived experiences, and prior knowledge; teachers are the sole knowers in the classroom while learners’ resources and strengths are rarely recognized and included in instructional and assessment practices.

3.1.2 Reliability

The Rasch personal reliability is 0.82 with a separation index of 2.14, suggesting that participants can be reliably differentiated into approximately three strata of enactment levels

Variable Map



F346HHH: Siti cares for and respects her students. She encourages students to be independent learners and to investigate and build understandings of their own, and she involves them in setting criteria and goals for their learning. She constructs her teaching practice to be engaging for all students and integrates a variety of assessment approaches into her teaching. Siti interacts with students to provide constructive feedback and adjusts her practice appropriately. She monitors and facilitates collaborative learning among her students.

F236MMM: Rebecca cooperates with some parents/community members and draws on some students' home culture as examples to design their learning experiences. Overall, she cares for and respects her students, although she sometimes engages in stereotypical thinking. Rebecca sometimes involves her students in designing a lesson or setting classroom rules. She often has students concentrate on their own work but from time to time she encourages collaboration among students. Rebecca's classroom is mostly inviting and safe for the students.

F456MMM: Steven generally has a positive sense of professional identity as a teacher. He considers the role of advocating policy/practice change to accommodate his students' needs as a marginal responsibility rather than a priority. Steven supports student learning and relies on standards and personal experiences with some inputs from students to identify learning priorities and teaching strategies. He usually designs assessments on his own and sometimes integrates them into his teaching to evaluate his own practice and to give feedback to students. Steven sometimes reflects on and adjusts his practice based on students' needs to better motivate their interests.

F456LLL: Daniel believes good teaching is carrying out a series of steps to ensure that students attain curriculum expectations for their grade/year levels. He relies heavily on standards or packaged materials to identify learning goals and instructional approaches. He determines how student learning would be evaluated and uses assessments to check whether students meet the mandatory academic standards. Daniel reviews his students' test results, sometimes altering his practice to boost their scores.

Fig. 1 Variable map

using the formula $(4 * \text{separation index} + 1) / 3$ (Wright and Masters 1982). The item reliability is 0.98 with a separation index of 7.80, indicating that the items are sufficiently spread to define the hierarchical continuum of the TEES construct.

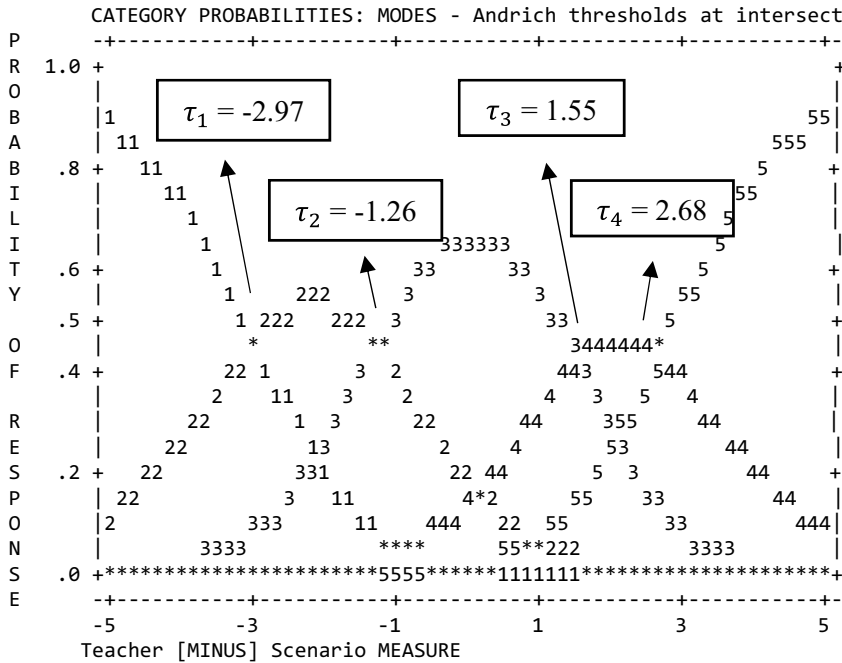


Fig. 2 Categorical characteristics curves

3.1.3 Categorical characteristics curves

Figure 2 shows the CCCs, suggesting that each response category has a unimodal shape, occupies a unique range not overlapped with another curve, and has approximately the same height (i.e., probability of response) as other curves. The Andrich thresholds (τ) suggest an increasing difficulty pattern of moving from one response option to the next, as intended. Specifically, when the difference between a person’s ability measure and an item’s difficulty estimate (X -axis in Fig. 2) exceeds -1.26 (τ_2) logit, participants are more likely to choose *About the same* than *Slightly lower*. As such a difference exceeds 1.55 (τ_3) logit when person ability is higher than item difficulty, participants are more likely to choose *Slightly higher* over *About the same*. Additionally, the average of the participants’ ability measures ($-1.14, -0.98, 0.06, 1.70,$ and 2.97) increased with each successive response category, suggesting that the response categories were used consistently across all the items. Overall, the response options function as intended.

3.1.4 Fit statistics

Table 3 presents item estimates and fit statistics. Most scenarios appear to have Infit- and Outfit-MNSQs within the range of 0.5 and 1.5, indicating productive measurement. The only misfit item is Scenario F236LLL, with an Infit-MNSQ of 2.13 and Outfit-MNSQ of 1.83 ($ZSTDs > 3.0$). Misfitting persons were also identified using the threshold of 1.5. Fifteen misfitting participants who gave unexpected responses to both misfit and nonmisfit items were identified. Further investigation into the 15 participants’

Table 3 Item Statistics

Scenario	Measure	S.E.	Infit MNSQ	Outfit MNSQ
F156HHH	1.96	0.18	1.05	1.05
F126HHH	1.45	0.18	0.82	0.83
F456HHH	1.26	0.18	0.66	0.66
F346HHH	1.31	0.18	0.90	0.88
F236HHH	0.96	0.18	0.91	0.92
F126MMM	1.87	0.18	1.01	1.03
F236MMM	0.20	0.18	0.60	0.59
F346MMM	0.21	0.18	0.53	0.52
F456MMM	-0.06	0.18	1.04	1.08
F156MMM	-0.77	0.17	0.70	0.69
F456LLL	-0.86	0.17	0.96	0.98
F346LLL	-1.27	0.17	1.02	1.19
F156LLL	-2.06	0.18	1.28	1.18
F236LLL	-2.13	0.18	2.13*	1.83*
F126LLL	-2.07	0.18	1.18	1.09

*ZSTD > 2.5

responses suggests that the six most misfitting participants (Outfit and Infit > 2.5) gave multiple unexpected responses to both fit and misfit items and that these participants repeatedly gave higher than expected responses to the higher level, more difficult items and lower than expected responses to the lower level, easier items. Of the six most misfitting participants, two had response times less than the median duration specified earlier. It is possible that these participants might not have carefully attended to the instructions explaining the comparative survey task. Examination of the remaining participants shows that these participants not only had fewer misfit responses, but the misfit responses seemed to be due to the matter of degree (i.e., choosing Much higher versus Slightly higher) rather than not using the responses as instructed.

Because the purpose of this study is to validate a scale consistent with the Rasch measurement principles, the six most misfitting respondents were removed according to the Rasch literature (Linacre 2010, 2016). The findings reported in this article are based on the trimmed sample of 72 teachers. Regarding the results of the untrimmed sample, the calibrated item location in the variable map was largely unchanged. The Rasch person and item reliability and separation, as well as response category function, were acceptable and improved after trimming. Scenarios F236LLL and F156LLL were misfit (MNSQs > 1.5; ZSTD > 3.0), and only Scenario F236LLL remained misfit after trimming.

3.1.5 Dimension analysis

The PCA results of item standardized residuals show that the eigenvalue of the first dimension is 3.3, suggesting that this component accounts for variance as strongly as the contribution of approximately three items. The correlations of residuals come from clusters of items measuring the lower and higher levels of equity-centered practice. The disattenuated correlation between the two clusters of standardized residuals is 0.3,

Table 4 Profile of interview participants

Pseudonym	TEES raw score	Subject	Level	Years of Teaching
Anup	56	social studies	secondary	> 10
Fatima	52	mathematics	secondary	> 10
Jason	48	geography	secondary	5~10
Evelyn	45	English	elementary	5~10
Seo-Ping	41	social studies	secondary	> 10

suggesting that the items most likely measure a single, rather than two distinct, dimension. This result was accepted with caution, and no separate analytical approach was attempted.

Altogether, the Rasch analysis result suggests that the survey data reported by Singapore teachers confirm the Rasch model expectation, which is grounded in the construct theories of equity-centered teaching practice. Overall, participants appeared to follow and apply the comparative survey instruction, and the TEES Scale scores can reliably differentiate lower- from higher-scoring participants.

3.2 Qualitative analysis

Table 4 presents interview participants' profiles. Participants' TEES scores range from 41 to 56, representing approximately one standard deviation below and above the mean of 47. I first discuss participants' survey taking experience, followed by the self-reported pedagogical practices among participants with higher and lower scores.

3.2.1 Survey experience

All the interview participants indicated that the survey instruction was clear and that they understood the comparative survey task. Nevertheless, three interviewees raised concerns over the judgment that one must exercise when comparing one's practice to the practice captured in the scenarios. Specifically, they pointed out that the frequency of enacting certain practices and the meaning of "lower than" or "higher than" could pose some challenges when selecting a response. Despite the concerns, when participants were asked to share their teaching practices that they reflected on and how they decided to select a specific option to a scenario, all the participants were able to elaborate on why they selected a specific response option, and their elaborations of the response processes were consistent with the design intention.

Regarding the scenarios, three participants explicitly stated that the scenarios were "interesting," "intriguing," and "rich in detail," allowing them to "connect," "resonate with," and "think and reflect" on their own practices. Across the five interview participants, detailed discussions about their teaching principles and practices were common. During the interviews, the participants expressed that they could resonate with a scenario, immediately signaled their disagreement with a scenario because the practice was too rigid or expressed aspirations to practices captured in a scenario. It appears that, because each scenario provides a rich picture of a teacher's practice, as opposed to the

conventional one-statement discrete item, participants can better relate to the descriptions. Although the scenarios are engaging and relatable, they were undoubtedly cognitively demanding. Two participants mentioned fatigue approximately halfway through and suggested shortening the length of the survey to 10 or 11 items.

3.2.2 Participants' self-reported pedagogical practices

In sharing their pedagogical practices, all participants mentioned the constraints and contextual factors that they encountered and worked with in a highly centralized, exam-oriented education system. The factors that participants commonly mentioned included: (a) constraints of time among competing goals, such as exam mandates and student learning; (b) the competitive nature and norms exacerbated by parent expectations and tuition; (c) standardized, prescribed, and formulaic exams, rubrics, and syllabus; (d) the leadership, school culture, and one's subject discipline training; and (e) the political sensitivity of some topics, such as racism. Despite the common boundaries, structures, and parameters that condition teachers' work, higher- and lower-scoring teachers diverge in their views of learners, knowledge and knowledge construction, perceived professional roles and identities, and instructional practices in significant and important manners. In addition to the differences, all participants also share characteristics that differ from the construct theories, as further elaborated below.

Teachers with higher TEES scores. Three participants -- Anup, Fatima, and Jason -- had higher TEES scores. They selected *About the same* to more difficult moderate-level scenarios, such as F236MMM, and high-level scenarios, such as F236HHH and F346HHH, and *Much higher* to all low-level scenarios. While they teach different subjects and are each unique in how they work with students, they share several characteristics in their self-described teaching practices.

Recognizing and including students' culture and lived experiences. First, these teachers shared how they recognized and attended to students' home culture, interests, needs, and experiences when designing and implementing instructional materials. For instance, Anup indicated that teachers are curriculum interpreters, designers, and implementers and that their priority is not to adhere closely to the policy but to consider "how it [curriculum] touches the kid in the classroom." As a social studies teacher, Anup shared how he looked for additional materials, such as cartoons, to cater to students' interests and modified the use of language to build their confidence and their vocabularies in discussing social issues. Similarly, Fatima, a mathematics teacher who often teaches NT students, discussed how she would look for additional resources, such as dance or music videos that demonstrate certain mathematic concepts and that are "related to real life," to motivate her students. Jason, a geography teacher, also elaborated on how he observed students' interests, needs, and prior knowledge, together with the contextual information of the neighborhoods where they come from in the first few weeks of the semester to adjust his instructional approaches.

Caring, demanding, and seeing students from an asset-based perspective. Recognizing their students' experiences and backgrounds, these teachers cared for and had high expectations of their students. They valued students' perspectives and worked to build students' capacity to be independent learners. They also appeared to have a more mutual and dialogical relationship with their students. For instance, Fatima discussed how she resonated with Scenario F126MMM, Timothy's practice, and particularly, Timothy's high expectations of his students:

I set high expectations for the students...when I first took them in Secondary 1, only three of them had actually ever passed math. Out of 40 of them, 37 have never passed math since Primary 5. When I told them that I'm setting expectations that 100%, all of them, will pass math at the end of their Secondary 1, they called me crazy. 'Cher, you crazy, ah? We never pass our math; then you think we can pass.' ... I said, 'Well, that's in the past. Yes, you may not have passed math, but don't you want to try this year to pass math?' And I told them, 'You can reach this goal if we are willing to work hard together.'

Fatima's "demand" for her students' academic success is connected to her strong belief in and care for them. Such a strength-based perspective of students is also observable in Anup's and Jason's sharing. For instance, Anup talked extensively about how he "respected the things students bring to the table," "encouraged them to buzz," and "worked with students and used some of the approaches they suggested in the class."

Demonstrating pedagogical flexibility. In relation to how these teachers view and work with their students, they demonstrated pedagogical flexibility — utilizing a broad variety of instructions relevant and meaningful to their students' interests and life experiences, using ongoing assessments to scaffold student learning, facilitating interactions and collaborations in the classroom, and offering a safe space for mistakes. For instance, Anup talked extensively about how he does not believe in "giving the answers" to the students and resists "offering the model essays" to them. He outright rejected Scenario F156LLL, Karen's practice, since she is "rigid" and a "transmitter of knowledge," and he discussed in detail how he usually would take students' perspectives and approaches, pose questions, and work with them to build their argument-making capacity.

Similarly, Jason discussed how he invited students to "co-construct" the learning target and criteria with him, facilitated students "bouncing ideas off each other, gathering different perspectives and sharing their opinions," and encouraged students to build understandings of their own, rather than seeking the right answers. Such pedagogical flexibility was also present when Fatima shared how she tailored her instructional approach:

For the better ones, I probably challenge them with maybe coming up with their own work problem ... with a real-life context. For the weaker ones, I'm not saying I don't challenge them. I do, but I scaffold ... I will give them some background to the scenario, and then after, that they build on it.

With a strong belief in her students, Fatima worked to better understand her students' needs and use different approaches to make learning relevant to their lives.

Demonstrating reflexivity and taking ownership to create changes within the system. Finally, all three teachers demonstrated the capacity to step back and critically reflect on their assumptions, biases, positions, and lenses and to take ownership of their work to make changes within the system. For instance, Jason shared that he "always tried to gather feedback from [his] students at the end to know what could be [his] blind spots and how receptive they are." Throughout the interview, Jason continued to reflect on how his instructional decisions could be based on his "one-sided assumptions and judgments" about his students and signal an area of improvement for him. Anup also reflected on Singapore's streaming policy and his role in the system:

Everything is so programmed ... I don't think it is a rocket science, when you can say, 'Oh, OK, when a child does like this, it's going to be like this at the end of four years' ... the kinds of projections that they make ... I do not know whether we

are trapped in a self-fulfilling prophecy or what ... I also realized that I'm part of a system ... I do voice my concerns ... there could be other factors that interfere. But then you know, this is the nature of the system. This is how it works.

While acknowledging the "system" that frames his work, Anup repeatedly discussed how he worked to resist the norms in his classroom. The ability to step back and reflect on the taken-for-granted practices and norms seems to allow these teachers to take agency and ownership to create changes.

Ultimately, these teachers' practices appear to be grounded in a commitment to teaching and learning that exceeds fulfilling the exam mandates. Moreover, these teachers' practices, which work against grain, continue to strengthen their conviction of what teaching and learning can be in a centralized and exam-oriented education system.

Teachers with lower scores. Evelyn and Seo-Ping, a primary English teacher who has never taught "low-progress" students and a secondary social studies teacher who often has NA/NT students, had lower TEES scores. They chose *About the same* to easier moderate-level scenarios, such as F456MMM, and low-level scenarios, such as F456LLL. Despite their different teaching experiences, they shared several characteristics in their self-described practices.

Emphasizing students' exam ability with little acknowledgment of students' lived experiences, culture, and prior knowledge. Throughout the interviews, both Evelyn and Seo-Ping focused mainly on the learning goals set by the rubrics, which were developed by their school leaders to align with the standard rubric from the Ministry of Education and the national exams. While Evelyn mentioned that students from expat families "had more interesting things and different perspectives to share than local students," there was no discussion around how students' lived experiences play a role in teaching and learning in her classes. Rather, Evelyn repeatedly mentioned how the exams and limited time render the practice described in the high-level scenarios "impossible." Additionally, referring to the exam-driven system, Seo-Ping outright rejected how students' culture is relevant to their learning in social studies subjects:

I don't see the home culture as very useful ... it doesn't really have that much to do with what I'm supposed to teach. I'm not teaching home economics...I talk about the experience when they [students] read the newspaper...as for traditional culture, it didn't seem that relevant.

Both teachers seemed to see culture as distinctive patterns of behaviors or lifestyles shared by a clearly bounded social group, rather than seeing culture as complex, multi-layered, and evolving ways of understanding and experiencing the world. As such, these teachers did not recognize the relevance of culture in teaching and learning.

Viewing students as receivers of knowledge and holding a deficit view toward "low-progress" students. In relation to these teachers' minimal recognition of students' lived experiences and prior knowledge, they tended to be hesitant about offering space and building students' capacity to become independent learners. For instance, Evelyn explained why she could not completely allow students build their own understanding:

I want them to be independent and to be able to know, like, when they see a certain theme, how do they go about deconstructing that theme ... ultimately, when they write whatever story it is according to that theme, there is a certain way whereby a theme needs to come out.

Similarly, Seo-Ping often described her students as needing “a lot of handholding” and “spoon feeding”; otherwise, they “would not know what to do if given space to set their goals.” She further rejected the importance of students being independent learners because students “have to first know exactly what they should know about within the syllabus for the national exams,” and “their learning has to be very structured.” Since the exams set the learning goal, students are deemed receivers of official, exam-oriented (and thus highly valued) school knowledge, rather than independent knowledge builders.

The teachers’ views of students, knowledge, and knowledge construction are closely connected to their views of the “low-progress” students. Both teachers talked about how these students are “not very motivated,” “less willing,” “not as bright,” and “have a lot of other emotional kinds of baggage that they bring with them.” As such, they need the “basics” and “lots of handholding.” With a static view of knowledge and an essential view of ability, both teachers tended to see students who do not do well on exams as inherently “lacking.”

Exhibiting limited pedagogical approaches and teaching to fulfill the exam mandates. Since both teachers frequently referred to the exam mandates, their pedagogical decisions and approaches became limited by the assessment rubrics, which were developed by school-level leaderships to mirror the exams. For instance, Evelyn indicated that she did not have much room to adjust her teaching approaches and did not use a variety of assessment approaches because the learning goal, rubrics, and her scheme of work are “pretty set,” and students in her “top class” have “already attuned their minds thinking about ‘how do I fulfill this set of criteria ... to hit as close as I can to a hundred.’” She further provided an example of the rubric that would specify certain ways of writing a “good and interesting” introduction and that writing outside of the rubric’s model approaches did not result in high scores. While acknowledging the somewhat limiting approach, Evelyn shared that, ultimately, teachers mark exams according to the rubrics, and students mostly write to fulfill the rubrics. Similarly, Seo-Ping repeatedly embraced the idea that a structured, tried-and-true approach to imparting knowledge to students to ensure that they do well on the exams is the primary role and responsibility of teachers.

These teachers believe that fulfilling the exam mandates is the primary goal of learning, although they acknowledge that learning is more than achieving high scores on exams. Either being constrained by or actively embracing the narrow view of learning, in these teachers’ sharing, the “system” largely determines their role as teachers and how they consider their work.

Shared characteristics among the participants. All interview participants, regardless of their scores, share two characteristics that diverge from the construct theories of equity-centered teaching practices initially conceptualized in mostly Western contexts. First, when asked about the aspect of collaborating with families and/or community members to understand students’ backgrounds and to inform instructional design, all of the teachers either explicitly indicated that they did not do so, or they talked about how they would contact the families when there were disciplinary or academic issues, when they needed to update parents, and/or when some students had special needs. That is, while the teachers do work with or engage the parents, collaborating closely with parents and tapping into student culture to design the learning experience seem less familiar among Singapore teachers. This aspect might explain why Scenario F126MMM, which has the explicit wording of “collaborating with parents and community members,” was more difficult than hypothesized.

Second, when asked about their roles as change agents to facilitate students’ capacity to enquire and take actions for issues of social justice or to advocate for policy changes, these

teachers gave ambivalent, mindful, and cautious responses. Most teachers talked about their roles as classroom teachers since policy changes are the purview of school leaders or occur at the ministry level. This finding is perhaps not surprising given Singapore's highly centralized education system. Additionally, when participants shared their ideas about addressing social justice issues, they mostly talked about "teaching the right values" or "imparting their own sense of social justice to the students." While two participants mentioned addressing issues related to inequality or inclusion/exclusion, a topic included in the state's syllabus, the aspects of examining structural inequalities, investigating root causes, and taking actions to address inequity or injustice were not evident. Considering the political volatility of topics concerning inequity and culture in Singapore, these responses might not be surprising.

3.3 Mixed methods comparison

Table 5 presents a joint display of the quantitative and qualitative results to address the mixed-methods question: To what extent do results from each strand converge or diverge. The first column lays out the *a priori* theories that inform the TEES construct from the lower to higher enactment levels. In the second column, the variable map presented earlier as Fig. 1 offers a good summary of the quantitative results, suggesting that the 15 calibrated items are sufficiently spread to measure a unidimensional construct of enacting equity-centered teaching practices from the lower to the higher levels as hypothesized in a reliable manner. Furthermore, individuals' scores can be interpreted to infer their differing classroom practices. The third column shows the summary of qualitative results, presenting the major patterns of practice drawn from the higher- and lower-scoring interview participants. Specifically, confirming their survey responses, higher scoring interviewees elaborated on how they recognized and purposefully tapped on students' lived experience, utilized a variety of pedagogical approaches to invite students in learning as a knowledge co-construction process, and demonstrated reflexivity and ability to resist and create changes within the system. On the other hand, lower-scoring interview participants generally complied with the narrow view of learning indoctrinated by an exam-oriented test-driven education system and emphasized students' role as knowledge receivers without recognizing the relevance of their lived experience in learning process. Overall, the quantitative and qualitative results confirm each other. The qualitative results offer further nuances and insights into the quantitative findings and the construct theories regarding how Singapore teachers enact varying levels of practice for equity in a highly centralized, exam-oriented education system.

4 Implications and conclusions

Given the persistent and endemic educational inequalities and increasingly diverse and yet politically divisive societies, teaching that is inclusive for all students with a commitment to recognizing and seeking ways to challenge systemic inequity is one approach to addressing persistent disparities. An instrument that can provide reliable and meaningful information about teachers' equity-centered practices and can be used to support teachers' professional learning is needed. This study used a mixed-methods convergent design to validate the TEES Scale, an existing instrument that was first conceptualized and developed in mostly Western contexts, among Singapore teachers.

Table 5 Joint display table

Joint Display Table

Theories of Equity-Centered Teaching Practice	TEES Scale		QUAN results		QUAL results
	Score	Item	Score Interpretations	Score Interpretations	Teachers' Self-described Practices
Higher-level enactment			Participants scoring between 52 and above are likely to select <i>About the same</i> to scenarios next to their locations, e.g., Scenario F346HHH.		- Recognize and include students' culture and lived experiences in teaching and learning
- Complex view of teaching	71	1	Siti cares for and respects her students. She encourages students to be independent learners and to investigate and build understandings of their own, and she involves them in setting criteria and goals for their learning. She constructs her teaching practice to be engaging for all students and integrates a variety of assessment approaches into her teaching. Siti interacts with students to provide constructive feedback and adjusts her practice appropriately. She monitors and facilitates collaborative learning among her students.	Participants scoring between 46 and 51 are likely to select <i>About the same</i> to e.g., Scenario F236MMM next to their locations and <i>Slightly higher</i> to all the scenarios below.	- Care, demand, and see students from an asset-based perspective
- Asset-based view of learners' lived experience and culture	71	1			
- A constructivist view of knowledge and knowledge construction	71	1			
- Teaching is value-laden and inherently political; teachers are agents of change	69	4			
	67	1			
	66	X			
	64	3			
	63	XX			
	61	X			
	60	X			
	59	X			
	58	2			
	57	XX			
	56	XX			
	55	XX			
	54	XXXX			
	53	XXX			
	52	X			
	51	XXXXXX			
	50	XXXXXX			
	49	XXXXXX			
	48	XXXXXX			
	47	XXXXXX			
	46	XXXXXX			
			Rebecca cooperates with some parents/community members and draws on some students' home culture as examples to design their learning experiences. Overall, she cares for and respects her students, although she sometimes engages in stereotypical thinking. Rebecca sometimes involves her students in designing a lesson or setting classroom rules. She often has students concentrate on their own work but from time to time she encourages collaboration among students. Rebecca's classroom is mostly inviting and safe for the students.		- Demonstrate pedagogical flexibility
			Participants scoring around 45 are likely to select <i>About the same</i> to Scenario F456MMM next to their locations and <i>Slightly higher</i> to all the scenarios below		- Demonstrate reflexivity and take ownership to create changes within the system
Lower-level enactment	45	0			- Emphasize on students' exam ability with little acknowledgement of students' lived experiences, culture, and prior knowledge in teaching and learning
- Technical view of teaching	41	XXXXXX	Steven generally has a positive sense of professional identity as a teacher. He considers the role of advocating policy/practice change to accommodate his students' needs as a marginal responsibility rather than a priority. Steven supports student learning and relies on standards and personal experiences with some inputs from students to identify learning priorities and teaching strategies. He usually designs assessments on his own and sometimes integrates them into his teaching to evaluate his own practice and to give feedback to students. Steven sometimes reflects on and adjusts his practice based on students' needs to better motivate their interests.	Participants scoring between 30 and 44 are likely to select <i>Slightly lower</i> to e.g., Scenario F456LLL next to their locations and <i>About the same</i> to all the scenarios below.	- View students as receivers of knowledge and hold a deficit view towards "low-progress" students
- Deficit lens of learners' lived experience and culture	41	XXXXXX			
- A static view of knowledge and knowledge construction	40	X			
- Teaching is value-free, and a teacher's responsibility excludes advocacy	39	X			
	38	-1			
	37	X			
	36	X			
	35	X			
	34	X			
	33	X			
	32	X			
	31	-2			
	30	X			
	29	X			
	28	X			
	27	X			
	26	X			
	25	X			
	24	X			
	23	X			
	22	X			
	21	X			
	20	X			
	19	X			
			Daniel believes good teaching is carrying out a series of steps to ensure that students attain curriculum expectations for their grade/year levels. He relies heavily on standards or packaged materials to identify learning goals and instructional approaches. He determines how student learning would be evaluated and uses assessments to check whether students meet the mandatory academic standards. Daniel reviews his students' test results, sometimes altering his practice to boost their scores.		- Exhibit limited pedagogical flexibility and teach to fulfill the exam mandates

The quantitative results suggest that the data largely confirm the hypothesized Rasch model. The 15-item TEES Scale measures the unidimensional construct of enacting equity-centered teaching practices from the lower to the higher level, can reliably differentiate participants into two to three enactment levels, and provides a qualitative interpretation of individuals' scores in relation to the construct. The qualitative results provide contextualized information about participants' survey experiences and their patterns of practices. Specifically, high-scoring participants tended to demonstrate a strong commitment to learning and teaching beyond fulfilling the exam mandates, to recognize and include the assets brought by students in their learning experiences, to care for and have high expectations of students, to have more dialogical relationships with learners, and to demonstrate pedagogical flexibility, reflexivity, and agency. Conversely, low-scoring participants tended to deem fulfilling the exam mandates as the goal of learning, perceived students as mostly knowledge receivers and those who performed less well on exams as "lacking," and demonstrated limited pedagogical approaches and lack of

agency in their work. In addition, the qualitative results revealed how Singapore teachers perceive some aspects of equity-centered teaching differently from the construct theories. Specifically, the participants tended to see the aspect of collaborating with parents to better understand and tap into students' experiences in designing instructional activities as difficult and even unfamiliar concepts. Moreover, Singapore teachers tended to have ambivalent, cautious, and sometimes limited responses to the aspects of teaching that challenge inequitable norms and rules and take actions to correct injustices, resonating with the findings of previous studies (e.g., Ro 2020). Importantly, these findings also offer supportive evidence for the unexpected results from the quantitative analysis. Overall, the qualitative results, which offer contextualized and thick descriptions, confirm and strengthen the quantitative findings.

This study has several limitations. First, the small and voluntary sample had more teachers teaching languages and humanities subjects, more teaching the secondary level, and more being in service longer. Additionally, approximately one third of participants were attending master's-level program coursework as part of their professional development. Since teachers' background and experiences can influence their survey responses, this study's findings could be skewed, and the generalizability is limited. Second, given the small sample size, it was not possible to obtain further validity evidence, such as the extent to which items exhibited biases against certain subgroups, or to evaluate the invariance of the scale structure across occasions. Third, this study relied on one-off interviews to understand participants' classroom practices without the opportunity for follow-up through a series of interviews or to observe teachers' practices over a prolonged period of time. Since part of this study's goal is to understand whether and how the scale scores can be interpreted appropriately to infer teachers' classroom practices, the current data collection approach provides a limited portrait of the phenomenon at hand.

Since validation is a continuous endeavor, more efforts are needed to provide further evidence supporting the proposed interpretations and uses of the scale across time, occasions, and contexts. Along a similar line, studies could investigate how the TEES Scale might be used to understand and support teachers' work with a commitment to promoting equity, justice, and cultural pluralism in other Asian countries that share similar sociopolitical contexts and centralized education systems and have an increasingly diverse school population due to regional migration. Moreover, the TEES Scale could provide further opportunities for research on teaching and teacher learning in Singapore. For instance, future studies could investigate how and under what conditions teachers understand and are supported in enacting equity-centered practices and how teachers' enactment of practices for equity shapes and is shaped by broadly defined student learning outcomes. Future studies could also explore how teachers' understandings of culture, diversity, and equity influence their curriculum design. In an increasingly diverse and uncertain world in which inequality is persistent, complex, and endemic, this study offers one possibility of supporting educators in considering the complexity of teaching and learning with a commitment to equity, justice, and cultural pluralism in their local contexts.

Acknowledgements I would like to thank Drs. Heng Tang Tang and Mardiana Bte Abu Bakar for their constructive feedback on the early version of the TEES Scale and ongoing support for this work during the implementation stage.

Funding This work was supported by Singapore's Ministry of Education's Education Research Funding Programme Planning Grant under OER PG03/19CWC.

Declarations The author has no relevant financial or non-financial interests to disclose. This manuscript is not under review by another publication.

References

- Abu Bakar, M.: 30 years of streaming: are we still giving some of our students half a loaf? *Mendaki Policy Digest* **1**, 123–138 (2009)
- Akiba, M., LeTendre, G.K. (eds.): *The International Handbook of Teacher Quality and Policy*. Routledge (2018)
- Alviar-Martin, T., Ho, L.-C.: “So, where do they fit in?” Teachers’ perspectives of multi-cultural education and diversity in Singapore. *Teach. Teacher Edu.* **27**, 127–135 (2011). <https://doi.org/10.1016/j.tate.2010.07.009>
- American Educational Research Association, the American psychological association, the National council on measurement in education: *Standards for educational and psychological testing*. (2014). https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- Anderson, K.T.: The discursive construction of lower-tracked students: Ideologies of meritocracy and the politics of education. *Educ. Policy Anal. Archives* **23**, 110 (2015). <https://doi.org/10.14507/epaa.v23.2141>
- Andrich, D.: A rating formulation for ordered response categories. *Psychometrika* **43**, 561–573 (1978). <https://doi.org/10.1007/BF02293814>
- Biesta, G.: Pragmatism and the philosophical foundations of mixed methods research. In: Tashakkori, A., Teddlie, C. (eds.) *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, 1st ed., pp. 95–118. Sage (2010)
- Bokhorst-Heng, W.D.: Multiculturalism’s narratives in Singapore and Canada Exploring a model for comparative multiculturalism and multicultural education: *J. Curriculum Stud.* **39**(6), 629–658 (2007). <https://doi.org/10.1080/00220270701506324>
- Borg, I., Shye, S.: *Facet theory*. Sage (1995)
- Buchanan, E.M., Scofield, J.E.: Methods to detect low quality data and its implication for psychological research. *Behav. Res. Methods* **50**, 2586–2596 (2018). <https://doi.org/10.3758/s13428-018-1035-6>
- Chang, W.C., Ludlow, L. H., Grudnoff, L., Ell, F., Haigh, M., Hill, M., & Cochran-Smith, M.: Measuring the complexity of teaching practice for equity: Development of a scenario-format scale. *Teaching and Teacher Education*, 82, 69–85. (2019).
- Chang, W.C: Measuring the complexity of equity-centered teaching practice: Enhancement and validation of a Rasch/Guttman scenario scale. *J Appl Measurement*, **22**(1/2), 35–59. (2021).
- Chang, W.C., Cochran-Smith, M.: Learning to teach for equity, social justice, and diversity: Do the measures measure up? *Journal of Teacher Education*.(2022). <https://doi.org/10.1177/00224871221075284>
- Chua, B.H.: Multiculturalism in Singapore: An instrument of social control. *Race & Class* **44**(3), 58–77 (2003). <https://doi.org/10.1177/0306396803044003025>
- Cochran-Smith, M., Ell, F., Grudnoff, L., Haigh, M., Hill, M., Ludlow, L.: Initial teacher education: what does it take to put equity at the center? *Teach. Teacher Edu.* **57**, 67–78 (2016). <https://doi.org/10.1016/j.tate.2016.03.006>
- Creswell, J.W., Clark, P.: *Designing and conducting mixed methods research*, 3rd edn. Sage, Los Angeles (2018)
- Goodwin, A.L., Low, E.L.: Rethinking conceptualisations of teacher quality in Singapore and Hong Kong: a comparative analysis. *Eur. J. Teacher Edu.* **44**(3), 365–382 (2021). <https://doi.org/10.1080/02619768.2021.1913117>
- Gopinathan, S.: *Education*. Singapore Press Holdings (2015)
- Grudnoff, L., Haigh, M., Hill, M., Cochran-Smith, M., Ell, F., Ludlow, L.: Teaching for equity: Insights from international evidence with implications for a teacher education curriculum. *Curriculum J* **28**(3), 1–22 (2017). <https://doi.org/10.1080/09585176.2017.1292934>
- Guba, E.G.: ERIC/ECTJ annual review paper: criteria for assessing the trustworthiness of naturalistic inquiries. *Edu. Commun. Technol* **29**(2), 75–91 (1981)
- Guttman, R., Greenbaum, C.W.: Facet theory: its development and current status. *Eurp. Psychol.* **3**(1), 13–36 (1998). <https://doi.org/10.1027/1016-9040.3.1.13>
- Guttman, L.: Introduction to facet design and analysis. In: *proceedings of the fifteenth international congress of psychology*, pp. 130–132. North Holland: (1959)

- Heng, M.A., Atencio, M.: 'I assume they don't think!': teachers' perceptions of normal technical students in Singapore. *Curriculum. J.* **28**(2), 212–230 (2017). <https://doi.org/10.1080/09585176.2016.1181558>
- Heng, T.T., Lim, L.: Diversity, difference, equity: how student differences are socially constructed in Singapore. *Cambridge J. Edu.* (2021). <https://doi.org/10.1080/0305764X.2021.1910204>
- Ho, L.-C.: "Our students do not get that equal chance": Teachers' perspectives of meritocracy. *Camb. J. Educ.* **51**(2), 173–193 (2021). <https://doi.org/10.1080/0305764X.2020.1796927>
- Ho, L.-C., Alviar-Martin, T., Leviste, E.N.P.: "There is space, and there are limits": The challenge of teaching controversial topics in an illiberal democracy. *Teachers Coll. Record* **116**, 1–28 (2014). <https://doi.org/10.1177/016146811411600402>
- Johnson, R.B., Onwuegbuzie, A.J., Turner, L.A.: Toward a definition of mixed methods research. *J. Mixed Methods Res* **1**, 112–133 (2007). <https://doi.org/10.1177/1558689806298224>
- Kane, M.T.: Validation. In: Brennan, R.L. (ed.) *Educational measurement*, 4th edn., pp. 17–64. The National Council on Measurement in Education & the American Council on Education (2006)
- Kwek, D., Miller, R., Manzoni, M.: "Bridges and Ladders": The paradox of equity in excellence in Singapore schools. In: Teng, S.S., Mazon, M., Kwek, D. (eds.) *Equity in excellence: Experiences of East-Asian high-performing education systems*, pp. 87–108. Springer, Singapore (2019)
- Lim, L.: Meritocracy, elitism, and egalitarianism: a preliminary and provisional assessment of Singapore's primary education review. *Asia Pac. J. Educ* **33**(1), 1–14 (2013). <https://doi.org/10.1080/02188791.2012.711294>
- Lim, L., Tan, M.: Culture, pedagogy and equity in a meritocratic education system: teachers' work and the politics of culture in Singapore. *Curriculum Inq* **48**(2), 184–202 (2018). <https://doi.org/10.1080/03626784.2018.1435974>
- Linacre, J.M.: Sample size and item calibration stability. *Rasch Meas. Trans* **7**(4), 328 (1994)
- Linacre, J.M.: What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*. **16**(2): 878(2002)
- Linacre, J.M.: When to stop removing items and persons in Rasch misfit analysis? *Rasch Measurement Transactions*. **23**(4), 1241(2010)
- Linacre, J.M.: *Winsteps® Rasch measurement computer program User's Guide* Beaverton, O.: Winsteps.com (2016)
- Linacre, M.: *WINSTEPS (Version 5.1.7) [Computer software]*. (2021). <https://www.winsteps.com/>
- Ludlow, L.H., Haley, S.M.: Rasch model logits: interpretation, use, and transformation. *Edu. Psychol. Measure.* **55**(6), 967–975 (1995). <https://doi.org/10.1177/0013164495055006005>
- Ludlow, L.H., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., James, J.B.: Measuring engagement in later life activities: Rasch-based scenario scales for work, caregiving, informal helping, and volunteering. *Meas. Evaluation Couns. Dev* **47**(2), 127–149 (2014). <https://doi.org/10.1177/0748175614522273>
- Ludlow, L. H., Baez-Cruz, M., Chang, W.C., Reynolds, K.: Rasch/Guttman Scenario (RGS) scales: A methodological framework. *Journal of Applied Measurement*, **21**(4), 361–378. (2020).
- Ludlow, L. H., Reynolds, K., Baez-Cruz, M., & Chang, W.C.: Enhancing the interpretation of scores through Rasch-based scenario-style items. In U. Luhanga & G. Harbaugh (Eds.), *Basic elements of survey research in education* 673-718. (2021). Information Age Publishing.
- Maul, A.: Moving beyond traditional methods of survey validation. *Measure.: Interdisciplin. Res. Perspectiv.* **15**(2), 103–109 (2017). <https://doi.org/10.1080/15366367.2017.1369786>
- Miles, M.B., Huberman, A.M., Saldana, J.: *Qualitative Data Analysis: A Methods Sourcebook*. SAGE Publications (2013)
- Morgan, D.L.: Paradigms lost and pragmatism regained: methodological implications of combining qualitative and quantitative methods. *J. Mixed Methods Res.* **1**(1), 48–76 (2007). <https://doi.org/10.1177/2345678906292462>
- Rasch, G.: *Probabilistic models for some intelligence and attainment tests* (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword: and afterword by B. D. Wright. The University of Chicago Press(1960/1980)
- Ro, J.: Curriculum, standards, and professionalization: the policy discourse on teacher professionalism in Singapore. *Teach. Teacher Educ.* **91**, 1–10 (2020). <https://doi.org/10.1016/j.tate.2020.103056>
- Talib, N., Fitzgerald, R.: Inequality as meritocracy: the use of the metaphor of diversity and the value of inequality in Singapore's meritocratic education system. *Crit. Discourse Stud.* **12**, 445–462 (2015). <https://doi.org/10.1080/17405904.2015.1034740>
- Tan, C.: Narrowing the gap: The educational achievements of the Malay community in Singapore. *Intercultural Educ* **18**(1), 53–64 (2007). <https://doi.org/10.1080/14675980601143710>
- Tan, K.P.: Meritocracy and elitism in a global city: ideological shifts in Singapore. *Int. Political Sci. Rev.* **29**(1), 7–27 (2008). <https://doi.org/10.1177/0192512107083445>

- Tan, J.: Equity and meritocracy in Singapore. In: Teng, S.S., Mazon, M., Kwek, D. (eds.) *Equity in excellence: Experiences of East-Asian high-performing education systems*, pp. 111–126. Springer, Singapore (2019)
- Tan, J.: Closing the Achievement Gap in Singapore. In: Clark, J.V. (ed.) *Closing the Achievement Gap from an International Perspective: Transforming STEM for Effective Education*, pp. 251–262. Springer (2014)
- UNESCO [United Nations Educational, Scientific, and Cultural Organization]: *Education for all 2000–2015: Achievements and challenges*. (2015). <http://unesdoc.unesco.org/images/0023/002322/232205e.pdf>
- VERBI Software: MAXQDA (Version 2020) [computer software]. VERBI Software. maxqda.com: (2019)
- Wolfe, E.W., Smith, E.V., Jr.: Instrument development tools and activities for measure validation: using Rasch models: Part II—validation activities. *J. Appl. Measure* **8**(2), 204–234 (2007)
- Wright, B.D., Masters, G.N.: *Rating scale analysis*. MESA Press, Chicago (1982)
- Zhu, D., Carterette, B.: An analysis of assessor behavior in crowdsourced preference judgments. In: Lease, M., Carvalho, V., Yilmaz, E. (eds.) *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation*. Geneva, Switzerland (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.