A peer reviewed, open-access electronic journal: ISSN 1531-7714

Participatory Construct Mapping

Fernanda Gándara, *Room to Read*

Abstract: Culturally responsive assessment frameworks are essential for enhancing fairness in educational testing and ensuring relevance across diverse settings. This article explores Participatory Construct Mapping as a practical method for improving the cultural and contextual validity of assessments. Drawing from feminist and decolonial methodologies, Participatory Construct Mapping involves test-takers in defining constructs, emphasizing the importance of voice, epistemological diversity, and positionality. By integrating participants' perspectives, this methodology provides a transparent foundation for item and test development. This article highlights the application of Participatory Construct Mapping in revising Room to Read's Adolescent Life Skills Assessment (ALSA) across Nepal, Tanzania, and Bangladesh. Through scenario-based discussions with adolescent girls, constructs were mapped to reflect skills relevant at home, school, and community levels. The resulting construct maps informed the development of culturally relevant items. Pilot testing and qualitative feedback ensured both psychometric rigor and participant engagement. We conclude that Participatory Construct Mapping supports broader test development goals and that its applications extend beyond test design.

Keywords: Cultural relevance, ALSA, Measurement, Participatory methodologies, Justice oriented, Construct maps

Introduction

Across the world, educators and psychometricians are rethinking what it means to measure learning fairly. From justice-oriented antiracist testing in the United States (Randall 2021; Randall et al., 2022; Randall et al., 2024) to community-rooted assessments of agency in East Africa (e.g. Sidle & Oulo 2023), new frameworks are emerging to confront a shared problem: traditional assessments often reproduce hegemonic epistemologies and their harmful effects on marginalized populations. Whether framed as antiracist, culturally responsive, or justice-oriented, these movements share a common ambition, which is to challenge

these dominant ways of knowing and to design assessments that do not erase, distort, or punish the knowledge of marginalized communities.

At the heart of this shift is context. Incorporating context into assessment has equity, pedagogical, cognitive and cultural foundations (Hood,1998; Lee, 1998; Qualls, 1998). Mislevy's (2018) sociocognitive approach defines context as the "linguistic, cultural, and substantive patterns" (LCS patterns) that shape both relevant and irrelevant aspects of assessments. Sociocognitive and culturally responsive approaches show that when tests ignore these patterns, they systematically disadvantage learners whose ways of knowing fall outside dominant traditions. Justice-oriented frameworks push this further, asking not only whether scores are generalizable, but whether they are just, whose knowledge counts, whose practices are valued, and whose realities are rendered invisible (Randall et al., 2022).

Justice-oriented antiracist frameworks also emphasize the importance of power and positionality. Randall et al. (2021) urge assessment professionals to engage in self-reflection and address positionality throughout the test development process. In addition, feminist and decolonial traditions remind us that knowledge itself is political: it is shaped by power, history, and whose voices are heard (Podems, 2010; Sielbeck-Bowen et al., 2002). If assessment is to account for context in just ways, it integrates the voices of those most often measured but least often listened to, treating their ways of knowing not as deviations, but as foundations for what counts as valid knowledge. Therefore, drawing from feminist and decolonial methodologies, I believe that a path towards effective justice-oriented testing is to focus more explicitly on the issue of *voice*, from the construct articulation phase, expanding the core epistemological foundations of assessments.

Rethinking Construct Articulation

Early stages of test development involve construct articulation. Constructs are defined as "the concept or characteristic that a test is designed to measure." (AERA et al., 2014, p 217) yet there are multiple ways to think about constructs and by extension, measurement. Four key paradigms are noted in this regard. From a *classical* perspective, constructs exist, and the role of measurement is to discover the empirical relationship between the constructs and other attributes (Michell, 1997). From an *operationalist* perspective, measurement is about assigning numbers to objects or events according to rules (Stevens, 1946). From a *representationalist* standpoint, the assignment of numbers to events needs to reflect hypothesized qualitative relations between objects, ensuring the empirical relations are represented within a numerical system (Markus & Borsbroom, 2013). Lastly, a *latent variable* perspective assumes that a psychological attribute with a given structure (a construct) exists and can be probabilistically linked to a set of observable variables that statistically depend on the unobserved construct (Markus & Borsbroom, 2013). These different measurement paradigms rest on distinct ontological assumptions about constructs, which cannot be overlooked.

The way we think about constructs and measurements is crucial for test development and validation, and construct articulation deserves far more attention than it typically receives. In this regard, Randall (2021) presents a valuable framework for decentering knowledge at the construct level. Her framework connects construct definition to purpose, positionality, people/places, power, process, and products/consequences. She explicitly calls for anti-racist construct development that addresses these domains and incorporates the voices of the community. I would argue that these communities should not only be included in the process but also have meaningful influence over it. Bottomline, we need to focus on <u>who</u> makes the decisions around constructs and their characteristics. No matter how committed we are to being anti-racist, anti-colonial, and aware of our positionality, we all have blind spots and cannot fully recreate others' experiences. As Mislevy (2018) notes, constructs are constrained by designers' understandings of the knowledge, skills, and dispositions being measured. If we are serious about decentering knowledge, we must implement practical and effective participatory approaches throughout the test development process.

Construct Maps. Construct maps, or variable maps, represent the assumed structure of a given construct (Engelhard, 2013; Wilson, 2004). They offer more precision than a general construct by depicting a latent variable that ranges from the lowest to the highest extremes, providing a qualitative sense of the differences across this range. While not all constructs can be visualized using unidimensional maps, when applicable, construct maps can support accurate measurement (Wilson, 2004). These maps serve as the conceptual foundation for item development, the categorization and scoring of responses, and the selection of psychometric models that link observed data back to the construct (Figure 1).

Figure 1. Example of a Construct Map (Earth System Science)[(From Liu, et al. (2024)]

Level	description
5 (High School Standard)	Students are able to translate the movements of the Earth and the Moon into a complete description of the solar system's motion and demonstrate it, explaining the geographic significance of Earth's Movements. Examples include: ·Day-night cycle ·Seasons ·Climate zones (five belts) ·Noon solar altitude
4 (Middle School Standard)	Students are able to coordinate the observable movements of celestial objects with their actual movements, explaining: ·The Earth both revolves around the Sun and rotates on its axis ·The Earth revolves around the Sun once per year ·The Earth rotates on its axis once per day, causing day-night alternation and the apparent movement of the Sun in the sky ·A basic explanation of how the Sun's direct rays and the variation in day length form seasons and climate zones ·Demonstration of the movement of the Sun's direct rays causing changes in day length, and summarizing the patterns of day length variation.[3] Common Error: Seasons are caused by the change in distance between the Earth and the Sun.
3	Students understand: ·The Earth revolves around the Sun ·The Earth rotates on its axis However, students have not integrated this knowledge with observable movements to form explanations and may not recognize that the Earth both rotates and revolves. Common Error: Nighttime gets dark because the Earth revolves around the Sun in a day.
2	Students recognize: ·The Sun appears to move across the sky daily ·Students may think that the Sun revolves around the Earth Common Errors: ·All movements in the sky are caused by the Earth's rotation on its axis ·The Sun revolves around the Earth ·Nighttime gets dark because the Sun revolves around the Earth in a day ·The Earth is the center of the universe
1	Students do not understand the systematic nature of celestial objects' appearance. Students may not recognize that the Earth is spherical. Common Error: The Sun is below the Earth at night.
0	No evidence or deviates from the expected path.

Construct maps enhance transparency in construct articulation by delineating qualitative distinctions that help target varying levels of difficulty, an essential step in test development. In essence, construct maps make explicit the assumptions about what is important to measure and where specific skills or attributes are expected to fall along a continuum. Within the Rasch measurement tradition, they serve to position both items and individuals on the same latent continuum (Engelhard, 2013; Wilson, 2004). Yet their utility extends well beyond any single psychometric paradigm. Construct maps provide a framework for empirically testing the validity of construct assumptions. For instance, if items designed to target lower levels of a construct consistently emerge as more difficult than those targeting higher levels, this misalignment signals a flawed assumption about what constitutes easier versus harder manifestations of the construct.

Moreover, construct maps are powerful tools for examining epistemological diversity and voice in assessment. They make visible *whose* knowledge and values are embedded in the construct definition. If the attributes I consider relevant are absent from the map, and thus from the resulting test, I am less likely to view the assessment as a legitimate representation of the underlying construct. In this way, construct maps (and the items derived from them) can clearly show whose perspectives are prioritized and whose are marginalized. Making construct articulation a transparent process can therefore support justice-oriented measurement practices, by inviting critical reflection on the underlying assumptions in ways that decenter knowledge

Purpose

Culturally relevant assessments and justice-oriented measurement frameworks must actively integrate the perspectives of test takers. This article introduces and explains the concept of *Participatory Construct Mapping* and the resulting *Participatory Construct Maps* (PCMs). I argue that Participatory Construct Mapping is an intuitive and transparent method for meaningfully involving test takers in the process of defining constructs. PCMs are particularly powerful for three reasons: (a) they compel test developers to make their assumptions about constructs explicit, (b) they enable the integration of diverse perspectives at the construct articulation stage, and (c) they provide a clear foundation for item development. In the following sections, I will outline how Participatory Construct Mapping has been used to develop life-skills assessments in multiple countries as part of my work with Room to Read[1], a non-profit organization supporting children in historically low-income communities worldwide. Specifically, I will describe how we first employed Participatory Construct Mapping to develop assessments in Nepal, Tanzania, and Bangladesh.

Room to Read's Girls' Education and Gender Equality Program

Room to Read, founded in 2000, is built on the belief that addressing illiteracy and gender inequality is essential for creating a better world. Through its Girls' Education and Gender Equality Program (GEP), Room to Read supports children, especially girls, in developing the life skills needed to succeed in secondary school and beyond. The GEP works in collaboration with local governments and communities to implement and refine its programs in public schools across eight countries. The program's long-term goal is to integrate these life skills into educational systems, enabling schools and communities to deliver positive outcomes at scale.

Recently, Room to Read's GEP underwent a systematic revision of its life skills framework and curriculum. Through a multi-layered process, the team developed a new framework focused on five key domains: collaboration, critical thinking, decision-making, leadership, and resilience (Spangler, 2024). These life skills are more closely aligned with well-established models, such as the Big Five taxonomy (John &

[1] https://www.roomtoread.org/

Srivasta, 1999). This refined focus on fewer, research-backed domains has created an opportunity to update the Adolescents Life Skills Assessment (ALSA), which is used to evaluate the program's effectiveness.

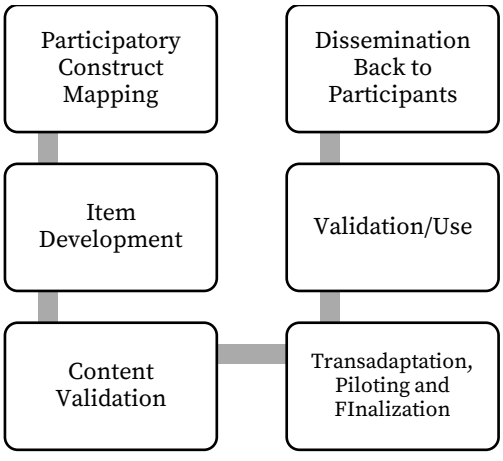
Room to Read's Adolescents Life Skills Assessment - ALSA

Room to Read's ALSA is designed to measure girls' life skills within the context of the organization's research and evaluation efforts. Its primary purpose is to assess whether students in the Girls' Education and Gender Equality Program (GEP) are developing key life skills at a faster rate than their peers. The original version of the ALSA, created in 2018 (Room to Read, 2022), has been used in seven out of eight countries where the program operates.

Since 2022, Room to Read's GEP Global Research, Monitoring, and Evaluation team has been revising the ALSA on a country-by-country basis. Reflecting changes to the program, the updated version now focuses on four key life skills - collaboration, decision-making, emotional resilience, and leadership - along with a complementary domain that measures gender knowledge and attitudes (Gándara, 2024). While the tool revision was triggered by changes to the programmatic framework, it also provided an opportunity to enhance the tool's cultural relevance and justice orientation. As a result, the ALSA revision focused on: (a) improving the cultural relevance of the tool, (b) creating opportunities for test takers to reflect on their own lives, and (c) incorporating context-appropriate items and assessment practices.

To meet these goals, the ALSA revision allowed the team to redefine its entire development process (see Figure 2). Instead of adapting a foreign tool to each country, Room to Read began developing tools from the ground up. The ALSA items would be developed to reflect participants' understanding about how each of the five domains manifested in their lives. Doing so would improve our capacity to measure culturally and contextually relevant impact. It would also expand the LCS patterns embedded in the tool, by incorporating the beliefs and practices of the populations being assessed. However, the ALSA would still need to yield comparable information across countries. Room to Read is a global organization, with a global curriculum, and being able to provide insights on comparative impact remains highly relevant. Thus, we decided to use a *mixed-item strategy*, and keep a core set of items that would be common across countries in each of the ALSAs, a.k.a. *global items*, as well as country specific items, a.k.a. *local items*. Each ALSA would include a core set of global items and the country specific local items.

Figure 2. Revised ALSA's Development Process



It was within this context that we adopted Participatory Construct Mapping as a technique. Our goal was to use these constructs to create local items from the ground up, thereby expanding the cultural and contextual relevance of the ALSA. Importantly, we wanted to make our assumptions explicit and, through

this process, offer a genuine opportunity to integrate participants' thoughts and experiences into the development of the ALSA.

Measurement Assumptions. We approach the development of ALSA with the assumption that each domain can be consistently observed through participants' knowledge, attitudes, behaviors, and beliefs. Evidence from multiple socio-emotional learning (SEL) and life skills assessments demonstrates that it is possible to reliably measure constructs such as decision-making, emotional resilience, and gender attitudes across diverse international contexts (e.g., OECD, 2021a; Pulerwitz & Barker, 2008; Soto et al., 2022). Moreover, we assume that the five ALSA domains manifest as attributes that are stable yet malleable, making them appropriate for tracking changes over time. This assumption aligns with research showing that SEL and life skills constructs, including those embedded in the ALSA, exhibit these properties (e.g., OECD, 2021b; Burrus et al., 2022). For instance, the OECD's Survey on SEL Skills targets constructs that are "malleable and susceptible to policy interventions" (OECD, 2021b, p. 18), which also map onto the Big Five framework. Our assumptions are thus supported by a growing body of research and measurement practice.

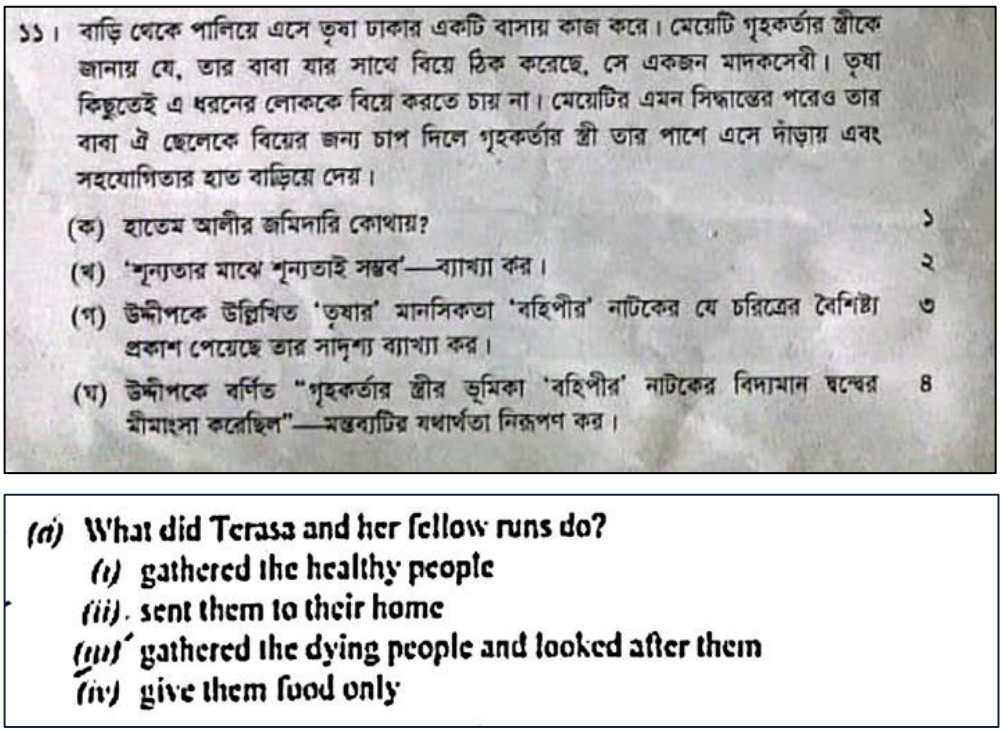
Guided by these assumptions, our team approaches test development trying to identify attitudes, behaviors, knowledge, and beliefs that serve as stable indicators of the five ALSA domains, yet remain responsive to positive change over time. Importantly, our teams prioritize attributes that are meaningful to the program participants and local stakeholders themselves. That is where Participatory Construct Mapping proves exceptionally powerful.

Measurement Constraints. In terms of logistical constraints, we aim to keep the administration time under 45 minutes. Based on historical analyses, we limit the ALSA to 50 items or fewer and prioritize local items. Additionally, we focus on constructing tests that can be effectively administered using paper and pencil. While the field of socio-emotional learning offers many innovative item types and psychometric tools, such as vignettes, forced-choice items and associated models (e.g., Bolt et al., 2014; Hontangas et al., 2015), these innovations may perform poorly in contexts where they aren't commonly used. Based on our experience testing forced-choice items, vignettes, negatively worded items, mirror drawing, listing exercises, and role-play scenarios and observations across eight countries over more than eight years, we concluded that such innovations generally do not enhance the overall quality of our measures (Room to Read, 2022). For example, none of the analyses allowed our teams to conclude that vignettes yielded more reliable scores, yet they were very time consuming and were thus discarded (see Room to Read, 2022 for more details). These results could be explained by the fact that innovative items tend to add cognitive load and may feel different to what students are used to in their educational systems, introducing construct-irrelevant variance. As a result, we try to rely on item types and formats that students have been exposed to in their curricula, that can meet our measurement goals, and that can be scored efficiently. As part of our development process, our country-level teams collect and examine tests used in the corresponding grade levels; we use this information to identify optimal item types and formats (see Figures 3 and 4).

Facilitating semi-structured conversations around a construct. Construct Maps are powerful artifacts that describe the different levels and attributes of a construct. These attributes should be both contextually relevant and psychometrically valuable, distinguishing between settings (e.g., school or home) and difficulty levels. Participatory Construct Maps (PCMs) should ensure that these attributes, and their meanings, are obtained as directly as possible from test takers. There are multiple methods for gathering input from participants, such as concept mapping (Allen, et al., 2017) which has been used to develop tools such as the AGAS (Sidle & Oulo, 2023), focus group discussions, and interviews. In the three experiences leading up to this article, we tried several approaches: focus group discussions, group interviews and semi-structured discussions. It became clear that focus group discussions provided good information about content but not about the perceived difficulty associated with the knowledge, behaviors, beliefs or attitudes named by participants. Group interviews (conducted mostly with adults) led to similar outcomes: a wealth

of unordered, and therefore psychometrically limited, information. Therefore, we chose to use semi-structured discussions to gather qualitative yet ordered data from participants

Figure 3. Example of Items from National Assessments in Bangladesh



১১। বাড়ি থেকে পালিয়ে এসে তৃষা ঢাকার একটি বাসায় কাজ করে। মেয়েটি গৃহকর্তার স্ত্রীকে জানায় যে, তার বাবা যার সাথে বিয়ে ঠিক করেছে, সে একজন মাদকসেবী। তৃষা কিছুতেই এ ধরনের লোককে বিয়ে করতে চায় না। মেয়েটির এমন সিদ্ধান্তের পরেও তার বাবা ঐ ছেলেকে বিয়ের জন্য চাপ দিলে গৃহকর্তার স্ত্রী তার পাশে এসে দাঁড়ায় এবং সহযোগিতার হাত বাড়িয়ে দেয়।

(ক) হাতেম আলীর জমিদারি কোথায়? ১

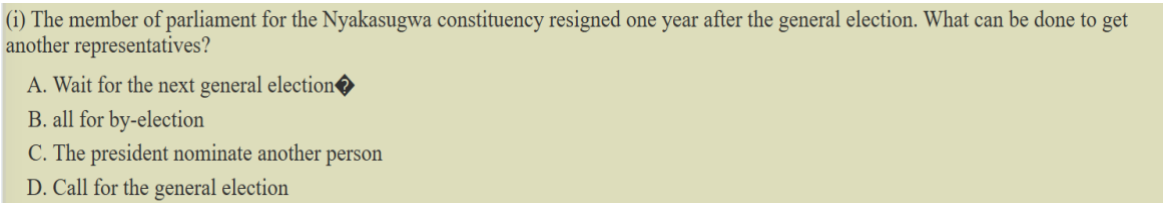
(খ) 'শূন্যতার মাঝে শূন্যতাই সম্ভব'—ব্যাখ্যা কর। ২

(গ) উদ্দীপকে উল্লিখিত 'তৃষার' মানসিকতা 'বহিপীর' নাটকের যে চরিত্রের বৈশিষ্ট্য প্রকাশ পেয়েছে তার সাদৃশ্য ব্যাখ্যা কর। ৩

(ঘ) উদ্দীপকে বর্ণিত "গৃহকর্তার স্ত্রীর ভূমিকা 'বহিপীর' নাটকের বিদ্যমান দ্বন্দ্বের মীমাংসা করেছিল"—মন্তব্যটির যথার্থতা নিরূপণ কর। ৪

(a) What did Terasa and her fellow runs do?
(i) gathered the healthy people
(ii) sent them to their home
(iii) gathered the dying people and looked after them
(iv) give them food only

Figure 4. Example of Item from National Assessment in Tanzania



(i) The member of parliament for the Nyakasugwa constituency resigned one year after the general election. What can be done to get another representatives?

A. Wait for the next general election
B. all for by-election
C. The president nominate another person
D. Call for the general election

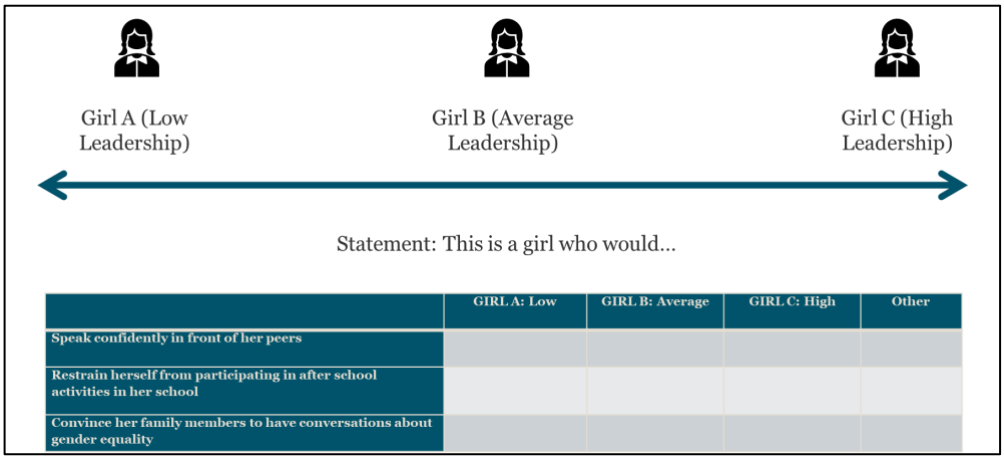
Using Participatory Construct Mapping to Revise and Strengthen the ALSA Developing Participatory Construct Maps

Semi-structured conversations may look different across contexts. In our case, we facilitated conversations by presenting a scenario involving three hypothetical girls: Girl A, with a low level of a given construct; Girl B, with an average level; and Girl C, with a high level (see Figure 5 for an example). We then read statements to participants, such as "*This is a girl who would speak confidently in front of her peers*," and ask them to identify whether the statement describes Girl A, B, or C. This simple, close-ended question allows the team to understand whether participants show variation in their responses and whether those responses are aligned with our expectations or not. After gathering everyone's responses, we ask participants why they thought the statement represented Girl A, B, or C, which typically leads to rich discussions about the topic. It is in this follow-up question, where the facilitators can dig deeper into assumptions, values, and ordered perceptions of the behaviors and ideas connected to the initial statement.

For example, one participant might say that "*speaking in front of peers*" is not a difficult skill unless it involves speaking in front of boys. Another participant might disagree, stating that "*girls are typically afraid of*

giving a wrong answer in front of their peers, even if they're girls." The facilitators use this information to guide the conversation further, asking follow-up questions such as, "*What is more difficult: speaking in front of male peers when you know the answer, or speaking in front of girls when you're unsure?"* The facilitators continue until they have enough information to qualitatively order the information for that theme. These discussions typically end with an open-ended question like "*What does [construct] mean to you?*" directed at each participant.

Figure 5. Example of a Scenario-Based Discussion for Leadership



Successfully facilitating these semi-structured discussions involves several layers of preparation. Consent and assent protocols must be in place, including permission to record the conversations. Extensive planning is required to develop effective statements and adjust them as the conversation unfolds. Additionally, discussions should be neither too long nor too crowded, as participants may lose interest. Facilitators must actively engage everyone – calling on every participant - and maximize the information gathered, while also exploring areas of disagreement, which often provide valuable insights about the construct. Initially, we had thought of keeping these conversations one hour long, but it quickly became apparent that 60 minutes was too long to hold participants' attention and interest. Across the three countries and among adolescents of different ages (typically 13-16 years old), the optimal time for these interviews was 45 minutes.

In Bangladesh[2], where we conducted several semi-structured conversations, we reached saturation after three semi-structured conversations covering the same construct. Saturation is typically defined as the point at which no additional data contribute to the category being developed (Glaser & Strauss, 1976). This finding is consistent with research showing that qualitative experiences such as focus groups yield most of their information in the first three occurrences (Guest et al., 2016). Research on focus groups also shows that as few as four individuals can provide accurate information on a given topic, and that six individuals can almost certainly identify trends shared by most of their social groups (Galvin, 2015; Guest et al., 2016). Like focus groups discussions, and similar experiences co-constructing assessments (e.g. Penuel et al., 2007), our conversations included 6 to 8 participants, which proved to be ideal for everyone to voice an opinion.

Due to safeguarding concerns, it is advisable to have at least two adults in the room at the time of the discussion. Another key aspect is ensuring that no one is present who might inhibit participants from freely expressing themselves; this may include teachers and even program officers, in our case. When addressing sensitive topics, child protection policies and referral mechanisms must be clear, available, and properly applied to mitigate any potential risks. And depending on the linguistic diversity of the region, translators may also be needed.

[2] And countries not explored in this article, such as India and Laos.

Transforming data from semi-structured conversations into construct maps. Developing Participatory Construct Maps (PCMs) for any construct involves conducting several semi-structured discussions with participants. After gathering sufficient information, the next step is to transcribe the conversations and conduct qualitative analysis. The goal of this qualitative analysis is to identify ordered attributes that can be mapped onto multiple levels. For the ALSA, this analysis focuses on identifying contextually relevant attributes across three settings: home, school, and community. By reviewing and reflecting on participants' opinions, we identify attributes that describe lower or higher levels of the construct. Many ALSA items require distinguishing attributes across four categories (a common format in the countries where we work), so our construct maps typically outline four distinct levels for each construct (see Table 1 for a generic template). An example of the initial version of the Leadership construct for the Bangladesh ALSA is provided in Table 2. This construct map, like all our construct maps, has been used for item generation and test development.

Table 1. Generic Template for ALSA Construct Maps

Construct, Country	**Description of skills**		
Levels	**Home**	**School**	**Community**
Level 4			
Level 3			
Level 2			
Level 1			

Developing PCMs forces us to make construct assumptions explicit. They also help us acknowledge the numerous interpretive and design decisions involved in articulating constructs and developing items to measure them. For example, in the case of *Leadership* in Bangladesh (Table 2), participating in sports and serving as a team captain emerged as examples of both school and community practices. Ultimately, we positioned this attribute under the community context, as it aligned more closely with extracurricular and community-based activities than with classroom experiences. Furthermore, the role of team captain was classified as representing Level 4, the highest possible level, after recognizing that, for these girls, becoming a captain symbolized the ultimate expression of leadership within sports.

Table 2. Pilot Construct Map for Leadership, Bangladesh

Leadership, Bangladesh	**Skills used to assert one's views, speak in groups and mobilize peers.**		
Levels	**Home**	**School**	**Community**
Level 4 Girls are fully engaged in leadership tasks; challenge norms; demonstrate	Girls can convince fathers to let their mothers work. They would argue that the mother should work not only because it is financially beneficial but also because it challenges restrictive gender roles.	Girls may speak loud in contextually appropriate situations. For example, a debate.	Girls participate in sports and may be captains in their sports' teams.

clear understanding of leadership in various contexts. Girls can express themselves freely. Girls are eloquent and can convince others to change their minds and/or engage in activities.	Girl can make her family members "understand" when there is a disagreement.	Girls can confidently speak in front of a big group, even in front of boys/men. Girls are used to talking to everyone at school. Girls may give speeches.	Girls participate in a variety of afterschool activities: acting, MC/anchoring, writing, sports, dance, sewing, drama, acting, planting trees and helping with classrooms. They do not stratify activities by gender and feel free to participate in all of them.
	Girls fearlessly confront family members in situation of gender inequality, such as witnessing their mother being mistreated.	Girls can solve problems in the classroom.	Girls are aware that community engagement and/or politics have historically been male dominated spaces. However, they do not feel intimidated in these spaces and believe that they can contribute to change those dynamics.
		Girls can get the captaincy in the classroom and talk to everyone. Being a captain means to keep the classroom clean, teach others the lesson, help bonding between other classmates, and generally assisting others in the classroom.	
	Girls are role models, speak clearly and without hesitation; possess extremely good communication skills.		
	Girls contribute to redefining societal expectations over the fear of negative social consequences.		
Level 3 Girls participate or encourage leadership but with some acceptance of gender norms;	Girls may confront family members in situation of gender inequality, such as witnessing their mother being mistreated, yet hesitate and may decide not to do it if the cost feels too high.		Girls participate in sports without aiming to become the captain.

they somewhat understand leadership roles. Girls can express some ideas with confidence. Girls want to influence people around them yet are limited by norms and self-beliefs.	Girls perceive convincing father to let mother work as demonstrating some leadership, but not high leadership. They see value in allowing the mother to work, primarily because of the financial benefits it would bring to the family. However, this is balanced by their concern about social repercussions.	Girls can sometimes give big speeches and talking in front of big crowds but would feel uncomfortable in front of men.	Girls participate in extracurricular activities but prefer "girl friendly" activities such as dancing.
		Girls can solve some problems in the classroom, but not all problems. They avoid more difficult circumstances, which may involve talking to people that they do not feel comfortable with.	Girls believe that community engagement and/or politics are male dominated spaces, yet they actively try to become part of them.
	Girls possess good communication skills and can speak without hesitation in many situations. However, they are not necessarily actively engaged in redefining gender norms.		
Level 2 Girls recognize leadership or participation in extracurriculars, but gender norms still limit their perspectives or actions. Girls cannot consistently express themselves in a confident manner. Girls have some interest but not ability in influencing people around them.	Girls think of convincing the father to let the mother work as an act of leadership, but they believe the risk of social backlash overshadows any potential gain. The fear that "people will say bad things about the family" reflects a hesitation to embrace leadership because of the social costs.	Girls speak with a soft voice in the classroom.	Girls are interested in sports but may not participate in them.
	Girls would like to but do not dare to start difficult conversations at home.	Girls value other people who have the courage to speak in big crowds.	Girls may participate in extracurricular activities but are hesitant to do so and are mostly trying to conform to the norms (rather than advancing their interest).
		Girls have some ideas on how to solve problems in the classroom but are	Girls believe that community engagement and/or politics are mostly male spaces but that there

		too shy or afraid to share them with others.	is room for women to participate. They do not try hard enough to engage in these activities.
	Girls may possess some communication skills but could be fearful of facing certain crowds, and do not try to change anything in the world around them.		
Level 1 Girls who show limited understanding or participation in leadership roles; are influenced by strict gender norms. Girls do not have the capacity to express what they think. Girls cannot influence people around them.	Girls think of convincing parents to let the mother work as an action that strongly defies societal norms, as a mistake, and would not engage in such discussion.	Girls do not speak up in the classroom.	Girls are not interested in sports.
	Girls do not challenge gender inequality in private settings.	Girls believe that giving speeches and/or speaking in front of big crowds is not impressive or relevant.	Girls do not have interest or feel it is appropriate to participate in extracurricular activities. This, because they do not want to be exposed.
		Girls would never be appointed captain of the classroom, as they cannot solve problems and/or talk to everyone.	Girls believe that community engagement and/or politics are male spaces. They do not have an interest in them.
	Girls lack the courage and the communication skills to engage in leadership tasks.		

Using Participatory Construct Maps for Item and Test Development

Construct maps are one of the building blocks used to develop items and assessments (Wilson, 2004). In our case, country level construct maps are used to develop *local items*, which make up 60% to 80% of ALSA items. As part of the ALSA development process, we draft approximately 100-150 items per country (including global items) and undergo extensive mixed methods piloting to create the final tools. These mixed methods include administering the items to students of different grade levels and social groups, and computing traditional item and test statistics, such as item difficulty, item discrimination, and internal consistency parameters. Quantitative analyses also focus on how item difficulty shifts with grade level and age, and whether item parameters remain similar across social groups. Our pilot process also includes five to six cognitive interviews per form and target group (e.g. girls of a given grade level), which focus on comprehension, relevance, comfort with the items, and suggestions for improvement.

The pilot process also focuses on participants' engagement and experiences taking the ALSA. For example, we ask participants who take the written tests about their level of comfort with the ALSA items and the reasons behind any potential discomfort, using closed- and open-ended questions. This feedback is triangulated with the information gathered through cognitive interviews and is integrated into the evidence map used to make final decisions about items. Most of the feedback provided by participants is neutral or positive. However, 13% of pilot participants in Bangladesh, 14% of participants in Tanzania and 20% of

pilot participants in Nepal, have identified at least one question that has made them uncomfortable, typically relating to bullying and/or gender violence. This information is exceptionally useful to develop assessments that gather meaningful data while minimizing discomfort.

Developing contextually relevant items that target key difficulty levels. Participatory construct maps offer a framework for developing culturally relevant items that target key levels of the constructs. For the ALSA, we target difficulty levels using a variety of item types. Since the adoption of PCM, we increasingly rely on multiple-choice items with four answer options, where each option corresponds to a specific level in the construct map. This item type, which ensures appropriate construct coverage, is essentially an ordered-multiple-choice (OMC) item (e.g., see Briggs et al., 2006), applied to contextually relevant life skills continua. OMC items resemble items used in many of the countries we work in (see Figures 3 and 4) and provide scoring benefits similar to those of traditional multiple-choice items. Furthermore, these items can provide diagnostic information (Briggs et al., 2006). OMC items enable teams to gain valuable insight into examinees' life skills, that can be used to improve our program design and implementation. While we use and value multiple item types (see Figure 7), adopting participatory construct maps amplifies the opportunities to target specific difficulty levels, and thus potentially improve psychometric properties of the tools, while enhancing contextual and cultural relevance.

Figure 6. From construct map to OMC pilot items – Example from Bangladesh

Context	
Levels	**Community**
Level 4	Girls are aware that community engagement and/or politics have historically been male dominated spaces. However, they do not feel intimidated in these spaces and believe that they can contribute to change those dynamics.
Level 3	Girls believe that community engagement and/or politics are male dominated spaces, yet they actively try to become part of them.
Level 2	Girls believe that community engagement and/or politics are mostly male spaces but that there is room for women to participate. They do not try hard enough to engage in these activities.
Level 1	Girls believe that community engagement and/or politics are male spaces. They do not have an interest in them.

Women engaged in community action and local politics:	a) Are wasting their time [LEVEL 1] b) Are generating backlash from community members [LEVEL 2-3] c) Are inspiring, but can not make a difference[LEVEL 2-3] d) Are transforming the world for the best [LEVEL 4]

Women engaged in community action and local politics:	a) Wasting their time [LEVEL 1] b) Creating problems [LEVEL 1] c) Unable to make a difference [LEVEL 1] d) Inspiring and making a difference [LEVEL 4]

A person like myself can become a community leader	Agree [LEVEL 3/4] / Disagree [LEVEL 1/2]

One of the challenges of using participatory construct maps to develop items, is identifying answer options that align with the intended construct levels. At the initial item-development stage, we assume a relationship between the answer options and specific levels of the construct. By creating and piloting a sufficiently large number of items covering similar topics and levels within the construct, we test these assumptions. Difficulty and discrimination parameters help us determine whether our assumed order relationships hold and whether answer options are differentiating between examinees along the continuum. These item parameters, and qualitative data, are used to identify well-performing items and to inform scoring decisions. For example, pilot data will suggest whether items – including OMC-type items – should be scored polytomously or dichotomously. Importantly, pilot data and analyses provide opportunities to test and refine order assumptions and the overall make-up of the constructs.

To date, we have developed 25 participatory construct maps in five countries[3]. We have used the technique to develop life skills assessments for girls and boys of different ages and backgrounds. In the three countries described in this article, we have gone through the entire test development process and finalized ALSA instruments. These tools are currently being used in program evaluation work. The adoption of

[3] The three countries described in this article, India and Laos.

Participatory Construct Mapping has expanded the foundations of test development and our ability to target culturally and contextually relevant abilities through meaningful items. We continue to explore item performance and examinees' experiences with the ALSA at the operational stage, to ensure that the broader goals of our assessment – like creating learning experiences for test takers - are fully met.

Figure 7. Example of Operational Leadership Items, Bangladesh, Grade 9

5. Women in politics:
   (a) Inspire me to become a leader
   (b) Disappoint me
   (c) Are irrelevant to me
   (d) Are creating numerous problems in our society

(E) Please indicate whether in the past 2 years, you have been a captain of any of these sports. (Tick or mark one option per activity)

Activity	Captain	
	Yes	No
Cricket		
Badminton		
Swimming		
Skipping Rope		
Ludo		
Frisbie		
Other		

29	It is very easy for me to win debate competitions at my school	Agree	Disagree

Discussion

For over two decades, researchers have emphasized the importance of addressing culture in assessments. As early as 1998, Stafford Hood argued that "*If assessment items and performance tasks are grounded in the cultural context of a particular group of color, those examinees may have a better chance of showing competence in certain learning areas*" (p185, Hoods, 1998a). Even then, scholars had already highlighted the need for assessment frameworks that reflected the linguistic, cultural, and substantive patterns of all examinees, including minorities and underserved populations. Similarly, critical methodologies, such as feminist approaches, have advocated for greater epistemological diversity in research, evaluation, and assessment. These theoretical frameworks share a common goal: amplifying marginalized voices to enhance fairness and justice in assessments. At Room to Read, our research, monitoring, and evaluation work is deeply committed to advancing this mission. As researchers and practitioners, we believe that advancing this mission requires developing practical approaches that can be refined and adopted by other organizations committed to the same goals.

In this article, I introduce Participatory Construct Mapping as an extension of traditional construct mapping in test development. Participatory Construct Mapping offers a different approach to co-creating assessments by expanding the range of perspectives involved in defining and representing constructs. This methodology compels test developers to make their assumptions explicit and transparent, creating opportunities to co-construct these assumptions in alignment with stakeholders' views, particularly those of examinees. Through in-depth qualitative work, such as structured face-to-face conversations with program participants, test takers, or equivalent groups, Participatory Construct Mapping broadens the epistemological foundations of test development. In doing so, it responds to calls from researchers and learners for

contextually relevant and participatory assessment practices (e.g. Solano-Flores et al., 2001; Suárez-Álvarez, et al., 2024).

The primary outcome of this work - Participatory Construct Maps (PCMs) - enhances culturally relevant test development by identifying ordered attributes of constructs that mirror participants lived experiences and epistemologies. This methodology also responds to current calls within psychometrics to learn directly from potential test users (Fremer, 2024) and to elevate the role of qualitative methods in measurement practice (Ho, 2024). Our experience using PCMs to revise the ALSA across multiple countries has been promising.

Participatory Construct Mapping aligns with justice-oriented and feminist methodologies, as it seeks to expand epistemological diversity and de-center dominant forms of knowledge. Participatory Construct Mapping can be adapted to diverse contexts. The defining feature of this approach lies in who contributes to construct articulation. While construct maps are typically used for item development, PCMs offer potential for validation work, an area that remains underexplored. The assumptions embedded in PCMs can be empirically tested through item types such as the ALSA's OMC-type items or similar formats that explicitly connect to construct levels. Furthermore, PCMs could be employed to evaluate how well constructs reflect the views of multiple subgroups, such as different ethnic or racial groups. By sharing and gathering feedback on PCMs, researchers can assess whether constructs are valued equally across subgroups and identify areas where important gaps persist.

Participatory Construct Maps could also be used to gather stronger validity evidence based on relationships with other variables. With sufficient reliability, scores could be generated by context (e.g., school-level decision making). Using PCMs in this way could add nuance to correlational analyses by allowing test users to examine whether domains interact at general levels (e.g. decision-making correlating with leadership) or at context-specific levels (e.g. decision-making at home correlating with leadership at school). Future research should explore the potential of PCMs to support validation. Finally, to fully gauge the promise of PCMs, future research should also examine their impact on other psychometric properties and across contexts with varying stakes.

Participatory Construct Mapping is not without its challenges. A notable risk is making incorrect order assumptions based on the information provided by participants. Pilot testing is critical to mitigate this risk. Additionally, the skills measured by the ALSA shift over time, meaning that PCMs require periodic revision. While PCMs can be used for sound item development, challenges at this stage remain. For example, item formats used in the ALSA, such as multiple-choice or OMC-type items, may be prevalent in educational contexts due to the legacy of colonialism and imperialism. Adopting Participatory Construct Mapping alone may not directly challenge or undo other problematic legacies embedded in educational systems and assessment practices. Lastly, the success of PCMs in serving minoritized communities depends on the ability to recruit key representatives from these groups and engage them in meaningful, honest dialogue, an aspect that can be particularly difficult to achieve.

Received: 1/6/2025. **Accepted:** 3/13/2026. **Published:** 3/25/2026.

Citation: Gándara, F. (2026). Participatory construct mapping. *Practical Assessment, Research, & Evaluation*, *31*(1)(8). Available online: https://doi.org/10.7275/pare.2837

Corresponding Author: Fernanda Gándara, Room to Read. Email: fernanda.gandara@roomtoread.org

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Bolt, D. M., Lu, Y., & Kim, J. S. (2014). Measurement and control of response styles using anchoring vignettes: a model-based approach. *Psychological methods*, *19*(4), 528.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. Educational Assessment, 11(1), 33–63.c https://doi.org/10.1207/s15326977ea1101_2

Burrus, J., Rikoon, S.H., & Brenneman, M.W. (Eds.). (2022). Assessing Competencies for Social and Emotional Learning: Conceptualization, Development, and Applications (1st ed.). Routledge. https://doi.org/10.4324/9781003102243

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

Fremer, J. (2024). Measurement reflections. *Educational Measurement: Issues and Practice.*

Gándara, F. (2024). Revising Room to Read's Adolescent Life Skills Assessment (ALSA). Available at https://www.roomtoread.org/the-latest/revising-room-to-reads-adolescent-life-skills-assessment/

Galvin, R. 2015. How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? Journal of Building Engineering 1:1–12.

Glaser, B., and A. Strauss. 1967. The Discovery of Grounded Theory: Strategies for Qualitative Research. New Brunswick, NJ: Aldine.

Guest, G., Namey, E., & McKenna, K. (2016). How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. *Field Methods*, *29*(1), 3-22. https://doi.org/10.1177/1525822X16639015 (Original work published 2017)

Ho, A. D. (2024). Measurement Must Be Qualitative, then Quantitative, then Qualitative Again. *Educational Measurement: Issues and Practice.*

Hontangas, P. M., De La Torre, J., Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, *39*(8), 598-612.

Hood, S. (1998). Culturally responsive performance-based assessment: conceptual and psychometric considerations. *Journal of Negro Education, 67(3)*, 187-196.

John, O. P. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research/Guilford.*

Lee, C. (1998). Culturally responsive pedagogy and performance-based assessment. *Journal of Negro Education*, 268-279.

Li, J., & Yu, J. (2024). Exploring students' understanding of "Earth's Movement" with OMC project. International Journal of New Developments in Education, 6(10).

Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning.* Routledge.

Mislevy, R. (2018). *Sociocognitive Foundations of Educational Measurement.* New York, USA: Taylor & Francis. Kindle Edition.

OECD (2021). Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills, OECD Publishing, Paris, https://doi.org/10.1787/92a11084-en.

OECD (2021). OECD Survey on Social and Emotional Skills: Technical Report, OECD Publishing, Paris. Available at https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/survey-on-social-and-emotional-skills/Technical%20Report%20SSES.pdf

Penuel,W. R., Roschelle, J., Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning, Volume 2 (1), 51-74*

Podems, D. R. (2010). Feminist Evaluation and Gender Approaches: There's a Difference? *Journal of Multidisciplinary Evaluation, Volume 6, Number 14* ISSN 1556-8180.

Pulerwitz, J., & Barker, G. (2008). *Gender-Equitable Men Scale (GEM)* [Database record]. APA PsycTests.

Qualls, A. (1998). Culturally responsive assessment: Development strategies and validity issues. *The Journal of Negro Education, 67 (3)*, 296-301.

Randall, J. (2021). "Color-neutral" is not a thing: Redefining construct definition and representation through a justice-oriented antiracist lens. *Educational Measurement Issues and Practice, 40(4)*, 82-90.

Randall, J., Poe, M., Oliveri, M., & Slomp, D. (2024). Justice-oriented, antiracist validation: Continuing to disrupt white supremacy in assessment practices. *Educational Assessment, 29(1)*, 1-20.

Randall, J., Slomp, D., Poe, M., & Oliveri, M. (2022). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. *Educational Assessment*, 1-9.

Room to Read (2022). Building the Adolescent Life Skills Assessment for Girls (ALSA for Girls). Synthesis Report on Lessons Learned. Available at https://www.roomtoread.org/media/tagpkajq/alsa-synthesis-report-mar-2022.pdf

Sidle, A., & Oulo, B. (2023). Assessment of a practitioner-derived framework for measuring girl's agency in East Africa. *Comparative Education Review, 67 (2).*

Sielbeck-Bowen, K. A., Brisolara, S., Seigart, D., Tischler, C., & Whitmore, E. (2002). Exploring feminist evaluation: The ground from which we rise. *New Directions for Evaluation*, *2002*(96), 3-8.

Solano-Flores, G. and Nelson-Barber, S. (2001), On the cultural validity of science assessments. J. Res. Sci. Teach., 38: 553-573. https://doi.org/10.1002/tea.1018

Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI.*Journal of Personality and Social Psychology, 123*(1), 192–222. https://doi.org/10.1037/pspp0000401

Spangler, A. (2024). Life Skills for Today. Room to Read. Available at https://www.roomtoread.org/the-latest/life-skills-for-today/

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677-680.

Suárez-Álvarez, J., Oliveri, M.E., Zenisky, A. L., & Sireci, S. G. (2024.) Five key actions for redesigning adult skills assessments from learners, employees, and educators. *Zeitschrift für Weiterbildungsforschung*, *47*, 321–343. https://doi.org/10.1007/s40955-024-00288-8

Wilson, M. (2004). *Constructing Measures: An Item Response Modeling Approach.* New York: Routledge.