Parental Involvement as a Holistic Concept Using Rasch/ Guttman Scenario Scales

Inna Antipkina[1] 🅭 and Larry H. Ludlow[2]

Abstract

Parental involvement (PI) in the education of their children is an important factor which should be taken into account when assessing and predicting children's school outcomes. However, PI encompasses numerous operationalizations from checking homework, to communication with school, to organizing cultural outings. This study describes a Rasch/Guttman scenario-based scale designed to provide a holistic approach to measuring the PI construct. The Parental Involvement SCenarios scale (PISC-9) was administered to 1,930 parents of primary school children from a sample representative of a Russian region. The scale has very good technical and construct validity characteristics. More specifically, raw scores on the PISC-9 may be represented as locations along a hierarchical continuum from relatively less to increasingly more time consuming and demanding parental behaviors.

Keywords

parental involvement, scenarios, vignettes, Rasch measurement, Guttman facet theory

Introduction

Parental involvement (PI) is a complex construct that has been defined as "parental participation in the educational processes and experiences of their children" (Jeynes, 2010); "any action taken by a parent that can theoretically be expected to improve student performance or behavior" (McNeal, 2014); and "manifestations of parents' commitments to their child's education affairs" (Bakker & Denessen, 2007). In practice, researchers choose very different aspects of parental activities to operationalize PI, a situation that leads to "inconsistent" definitions of PI in different studies (Wilder, 2014).

PI can be defined in different ways. Wilder (2014) lists operationalizations such as checking homework, home supervision, homework assistance, parent–child communication about schools, education expectations, participation in school activities, communication with school, reading with children, parenting style, and parental attitudes toward education. Pomerantz et al. (2007) add motivation, parental beliefs, and investments to the list. Bakker and Denessen (2007) write

[1]National Research University Higher School of Economics, Moscow, Russia
[2]Boston College, Lynch School of Education and Human Development, Department of Measurement, Evaluation, Statistics, and Assessment, Chestnut Hill, MA, USA

Corresponding Author:
Inna Antipkina, Institute of Education, National Research University Higher School of Economics, Potapovsky Pereulok, 16/10, Moscow 101000, Russia.
Email: iantipkina@hse.ru

about limiting TV, setting rules of discipline, encouraging sports, and organizing cultural outings. Attempts to systematize these definitions have divided PI practices into home-based versus school-based; controlling versus autonomy supportive; or naturally occurring PI behavior versus facilitated via parenting programs (Pomerantz et al., 2007). The need to include a PI measure in an educational study raises a question about which measure(s) of PI to choose from the range of aspects describing this loosely defined construct.

Measuring PI is associated with other specific challenges. First, PI is a culturally dependent construct related to the historical, demographic, political, and economic situations of people (Hornby, 2005). Second, PI may be prone to social desirability bias in respondent's answers. Studies have shown that most parents tend to rate themselves as involved or very involved, and systematic response bias in parents' answers "should be considered as a major problem" for the validity of measures (Bakker & Denessen, 2007). Previous suggestions about how to minimize this bias (e.g., triangulation involving asking parents, children, and teachers about PI) have not resolved the problem (Bakker & Denessen, 2007), and qualitative methods, such as in-depth interviews, are time-consuming and generally not appropriate for large-scale assessments.

Despite the fact that many manifestations of PI are of a different nature, they are all related to children's academic development. In a meta-synthesis, Wilder (2014) analyzed seven meta-analyses dedicated to relationships between PI and various academic outcomes of children. He concluded that the relationship between PI and academic achievement was positive, regardless of PI operationalization or measure of achievement (e.g., standardized test score, school grade, or teacher's rating score). These findings were consistent across grade levels and ethnic groups (Wilder, 2014).

The goal of this article is to present a unique scenario-based PI scale, developed in accordance with Rasch (1960/1980) measurement principles. The PI scale is part of the Russian international Performance Indicators in Primary Schools (iPIPS) project (Kardanova et al., 2018). iPIPS provides high quality actionable information about the development of children in primary school (Kardanova et al., 2018). To control for the context of children's development, iPIPS also includes questionnaires for teachers and parents.[1]

A Review of Published PI Scales

Despite the large number of studies dedicated to PI, few scales measuring PI in primary school have been published. We present a brief review of the three which were most useful in providing a context for our study (and according to Google Scholar they are among the ones most frequently cited: 937, 386, and 572 times, respectively).

The Family Involvement Questionnaire (FIQ; Fantuzzo et al., 2000) has a primary school version called the "Family Involvement Questionnaire-Elementary" (FIQ-E; Manz et al., 2004). The FIQ-E consists of 46 items for parents. The items address three factors: home–school communication, home-based involvement, and school-based involvement. Garbacz and Sheridan (2011) administered the FIQ-E in New Zealand, where the basic three-factor structure was confirmed but 13 items were inconsistent with the Manz et al. (2004) study.

The Alabama Parenting Questionnaire (APQ) is a free, publicly available instrument, aimed at children aged 6 to 18 years and their parents (Frick et al., 1999). It consists of 42 items and five subscales: (a) positive involvement with children (which includes items about school and home involvement); (b) supervision and monitoring; (c) use of positive discipline techniques; (d) consistency in the use of such discipline; and (e) use of corporal punishment. They reported an average Cronbach's alpha across subscales of .68.

In a study investigating different definitions of PI, Walker et al. (2005) administered 10 PI-related scales, including the "Parental role construction scale" (10 items, α = .8) and "Parents involvement in Home-based and School-based activities" (five items each, α of .85 and .82, respectively).

Finally, an early administration of the iPIPS attempted to measure PI using existing Likert-type items. As subscales 1 to 3 from the APQ had been translated into Chinese, Dutch, German, Spanish, and Norwegian, they were selected and translated into Russian for potential cross-cultural comparisons (Antipkina, Lyubitskaya & Nisskaya, 2018). A factor analysis of the items, however, led to the separation of home-based PI activities from school-based PI activities. Subsequent Rasch rating scale model analyses (Wright & Masters, 1982) of those two activity scales revealed: (a) low reliability for the school-based PI scale (Cronbach's α = .65); (b) poor goodness-of-fit for many items; (c) unidimensionality problems; (d) range restrictions; and (e) ceiling effects on the home-based PI, positive parenting, and monitoring scales. Because of the indistinct construct PI definition in existing instruments and the initial problems encountered in the iPIPS adaptation of the APQ scales, the iPIPS project sought an alternative approach for the measurement of PI.

Method

Rasch/Guttman Scenario (RGS) Measurement Approach

Our scenario approach to scale development includes an integration of Rasch measurement principles (Ludlow et al., 2014; Rasch, 1960/1980) and Guttman facet theory design (Borg & Shye, 1995). There are several general principles of Rasch measurement (Rasch, 1960/1980) that are used to guide the process of construct definition and scale development. These are rooted in the work of Thurstone (1928) and laid out explicitly in Ludlow et al. (2014). (a) Items should measure a single construct (unidimensionality). (b) Items should cover a wide range of content spanning the construct. Lack of variability may result in a failure to adequately capture the construct levels of all respondents. (c) Item difficulty levels should be evenly spread across the construct's range. For our purposes, we seek a "ladder-like" uniform continuum of scenarios that define the construct. (d) Different levels of the construct should follow a clear substantive hierarchical progression. Typically, this means we hypothesize some form of developmental continuum. (e) Items should all have the same relationship to the construct (equal discrimination). Equal discrimination (whether implicitly set at 1.0 or estimated as a constant across all items) distinguishes the Rasch model from other item response theory (IRT) models. Equal discrimination ensures the unambiguous interpretation of items at all levels of difficulty. Technically, this means that the item characteristic curves do not intersect (Engelhard, 2012). (f) An individual's response to one item should not depend on their response to other items (local independence). (g) The underlying hypothesized theory of the construct should be reflected in the empirical data collected.

A Rasch analysis may reveal that the data are not well-aligned with how the structure of the construct was hypothesized. In such a case, one must either reconceptualize the construct's theory or determine if there is something specific about the data (e.g., an inappropriate sample, confusing items, or social desirability bias) causing the mismatch between theory and data. In addition, unidimensionality, equal discrimination, and local independence are formal statistical assumptions subject to testing and numerous procedures exist for this purpose (for details see Andrich & Marais, 2019; Engelhard, 2012; Wilson, 2004; Wright & Masters, 1982). After scale construction, the Rasch psychometric analysis of the data serves as a confirmatory test of the extent to which these principles have been achieved and the assumptions met.

Guttman facet theory includes both design and analysis components but the RGS approach exclusively draws upon facet theory design because of its capacity to guide the systematic development of scenario items through the identification of a construct's "facets," which are the "concepts and contexts that guide empirical observations" (Borg & Shye, 1995, p. 13). Facets, in other words, are the different components of the underlying construct along which individuals vary.

The Rasch measurement principles provide the framework for building an instrument comprised of a hierarchical continuum of "lived-experience" scenarios. Guttman facet theory design facilitates the process of operationalizing the construct by breaking it into essential facets and then creating a "sentence map" template, which specifies the facets at various levels of intensity (Borg & Shye, 1995; Guttman & Greenbaum, 1998). The sentence mapping task provides the semantic mechanism to explicitly combine the different facet levels to create rich and authentic scenario items spreading across a wide domain of behaviors defining the construct. In essence, a sentence map illustrates how the facets may be woven together like the strands of a cable.

Scenario-based scales have been successfully developed to measure the productive engagement of older adults (Ludlow et al., 2014; Ludlow, Matz-Costa & Klein, 2019), teachers' enactment of practice for equity (Chang et al., 2019), readiness to return and participate in the community by psychiatric rehabilitation clients (Shen & Ludlow, 2018), living a life of meaning and purpose (Ludlow et al., 2019), and college faculty out-of-class availability (Reynolds, 2019). Greater detail in scenario scale development and the unique measurement challenges these scales pose may be found in Ludlow, Reynolds, Baez-Cruz & Chang (in press).

Development of the PI Scenario Scale

Positive significant correlations between different PI measures, for example, school-based involvement and home-based involvement (.4–.7), have been reported in many studies (Antipkina, Lyubitskaya, Nisskaya, 2018; Walker et al., 2005; Wilder, 2014). In Walker et al. (2005), for example, one conclusion was that "constructs do not operate in isolation," that is, "efforts to measure one construct may necessitate changing definitions and assessing others."

In the current study, we hypothesized that instead of decomposing the phenomenon of parental involvement into multiple subcomponents and measuring them separately it might be practical to approach PI as a single holistic, albeit complex, construct. Positive correlations between different measures of parental involvement (e.g., school-based and home-based activities) support this hypothesis, as well as empirical cases in which we see that parents often demonstrate a tendency toward a relatively constant range of involvement in different types of parental educational activities. We further hypothesized that holistic perspectives about the phenomenon of PI, depicted in scenario descriptions of parental involvement activities and beliefs (sometimes referred to as "vignettes"), would allow us to assess a broad range of levels of involvement, expressed through a common set of PI facets, which themselves could be expressed in different levels of intensity. These authentic lived-experience PI scenarios are then employed like traditional scale items for understanding different degrees and types of parental involvement in children's education.

Our PI scenario scale development process included the following steps:

Step 1: *Identification of essential facets based on previous theoretical and applied research*; "*facets*" *here are understood as critical components of the complex parental involvement construct.* The literature provides different classifications of PI activities such as home-based or school-based (e.g., Pomerantz et al., 2007); formal or informal (Lareau, 2011; Sénéchal & LeFevre, 2002); cognitive or noncognitive (Grolnick et al., 1997); and reading or math focused (Kastberg et al., 2013). Initially, we chose three facets: home-based activities, school-based involvement, and a focus on the child's well-being in academic settings. As a result of the cognitive lab feedback sessions (described in more detail below), however, the home-based facet was divided into parents differentiated learning activities at home versus family educational outings. Our final set of PI facets includes: (a) home-based, learning-related activities; (b) educational outings and extracurricular activities; (c) school-based involvement; and (d) time and effort invested in maintaining the child's well-being (how much parents take

into account the child's academic and extracurricular interests, and mental and health abilities and limitations when planning for the child's educational schedules and trajectories). Each of these four facets may be described as ranging from lesser to greater degrees of involvement, and they are all used together to define each scenario.

Step 2: *Development of meaningful narratives describing how each of these facets manifests in the behavior of people with higher, medium, and lower levels of the facet expressions.* During two cognitive lab sessions with two different groups of psychologists, parents, and teachers, we recorded their examples and illustrations of typical behaviors associated with parents who could be described as highly involved, moderately involved, and lesser involved in the education of their children. The function of these categorized facet narratives was to provide rich detailed qualitative descriptions of parental behaviors which then served as the foundation for the creation of scenarios capturing parental involvement levels. These narrative descriptions define what a parent "looks like" at different levels of parental involvement.

Step 3: *Creation of a sentence map template and scenario item specifications*. Our scenario development specifications included: (a) all scenario sentence structures followed the order of the facets presented in Step 1 and all facets were used in each scenario; (b) scenarios avoided as much as possible socio-economic-related descriptions, for example, references to parent's education or family resources; and (c) lower level scenarios, which are descriptions of relatively uninvolved parents, are nonjudgmental, for example, the wording is positive even when the scenario depicts a relatively undesirable level of parental involvement.

Table 1 presents the sentence mapping template used for constructing the scenarios. The columns contain the four facets, the rows contain the three facet levels. The cells of this template contain the lived-experience descriptions for a given facet written at each level of high, medium, or low PI. The facet level descriptions differ in the frequency of the activities and the quality of the activities the parent might engage in. The left-most column contains an ordinal scoring code that is assigned to each of the three facet levels. These codes are summed across the four facets to yield a scenario score. For example, a scenario written for the "high PI" level would consist of four facets, each written at the highest level corresponding to "code = 3" resulting in a "scenario score = 12." These scores then define our *a priori* expected order of difficulty for the scenarios. That is, a scenario scored as 12 should represent a holistic PI situation which is much harder for a parent to say that they are engaged in than a scenario with facets written at the "medium PI" level with a scenario score of 8. Furthermore, this hypothesized ordering of the scenarios represents our operational definition of the PI construct.

Step 4: *Development of scenarios*. Based on the mapping template and its narrative descriptions, nine scenario items were developed. "High PI" scenarios were written to be difficult for parents to achieve high scores; that is, parents would find it hard to rate their level of involvement higher than parent "X" depicted in a high level PI scenario. "Low PI" scenarios were written to be easier for parents to rate their level of involvement as higher than parent "X" in the scenario. Although scenarios could theoretically take any score value from 12 to 4 (through combining any of the possible combinations of four facets written at three levels), the higher level scenarios were written to range only from 12 to 11, medium scenarios were written to range from 8 to 7, and lower level scenarios were written to range from 5 to 4.

We anticipated that this restriction of the scenarios to initially just three levels of PI would create a clustering of the scenario difficulty estimates during the Rasch psychometric analysis (presented below in the discussion of the "variable map"). But, it is important to note that this decision to create three distinct levels of parental involvement scenarios was a deliberate strategy

Table 1. Template of Scenario Difficulties and Key Words.

Level and theoretical scores	Facet 1: home-based: learning-related PI	Facet 2: home-based: educational outings	Facet 3: school-based PI	Facet 4: focus on well-being
High PI: code = 3, maximum scenario score = 12	*Always/every day/very often* check homework, tutor, explain, engage the child in learning, demonstrate school subjects in practice, read, play educational games watch educational channels, experiment.	*Every week, very often* take the child to museums, libraries, exhibitions, workshops, science festivals, sight-seeing tours, historical places, zoo, enroll the child in extracurricular activities (sport, academic, arts).	*Always, systematically* visit school events, visit teacher–parent conferences, volunteer, help, organize, communicate with teachers, communicate with other parents.	*Always, usually* ask children about their interests, discuss school work-load and learning strategies, adjust extracurricular activities, monitor child's schedule, monitor child's physical condition and mental condition, discuss school-related things and problems.
Medium PI: code = 2, maximum scenario score = 8	*Often, regularly* help with homework, tutor, explain difficult subject topics, show interest in schooling, motivate.	*Sometimes—3 to 5 times a year* go to educational events (see above).	*Sometimes* visit key school events and meetings, *sometimes* communicate with teachers and other parents.	*Generally, usually, as a rule* ask child about his or her interests, discuss school things, reduce or enlarge work-load, stick to common school practices.
Low PI: code = 1, maximum scenario score = 4	*Always, usually* delegate learning things to child/teacher/school, teach child the family routine.	*Always, often, usually, as a rule, sometimes* go to playground, go to parks, watch TV, play computer, engage in other recreational activities.	*Sometimes, when strongly requested* go to school events/meetings, *always* receive information about school events from child or other parents.	*Always, usually, as a rule* follow teacher/school requirements, orient to typical/average school outcomes, orient to family tradition/opinion regarding learning goals and trajectories.

Note. PI = parental involvement.

to establish a proof-of-concept for this measurement approach. If meaningful and psychometrically sound scenarios could not be constructed for these three levels, then writing scenarios for the intermediate levels (i.e., 10, 9, and 6) would be fruitless. If successful, however, subsequent PI scale refinements could follow the strategy employed by Ludlow, Matz-Costa & Klein (2019) to reduce the clustering of the scenarios, leading to a more uniform distribution of the scenarios defining the PI construct's continuum.

The expected ordering of difficulty for the scenarios based on their scenario scores is shown in Table 2. Note how the scenarios written at score level 11 were constructed—obviously other combinations of facet levels were possible for generating a scenario at level 11.

Table 3 contains the 10 final scenarios: three each for the higher, medium, and lower levels of parental involvement and one "training" scenario explained in more detail below.

Table 2. Expected Order of the PI Scenario Continuum.

Scenario item	Theoretical score based on Table 1	Comments
Fedor	12	Maximally involved parent: (F1-3, F2-3, F3-3, F4-3)
Olga	11	Less than the maximum PI in Facet 2: (F1-3, F2-2, F3-3, F4-3)
Irina	11	Less than the maximum PI in Facet 1: (F1-2, F2-3, F3-3, F4-3)
Anton	8	Typical medium PI: (F1-2, F2-2, F3-2, F4-2)
Larisa	8	Typical medium PI: (F1-2, F2-2, F3-2, F4-2)
Tatiana	7	Lower than typical medium PI in Facet 1: (F1-1, F2-2, F3-2, F4-2)
Maria	5	Higher than typical lower PI in Facet 3: (F1-1, F2-1, F2-2, F1-1)
Lidia	4	Typical lower PI: (F1-1, F2-1, F3-1, F4-1)
Sergey	4	Typical lower PI: (F1-1, F2-1, F3-1, F4-1)

Note. F1 = home-based: learning-related PI; F2 = home-based: educational outings; F3 = school-based PI; F4 = focus on well-being. High level = 3; medium level = 2; low level = 1. F1-3 means Facet 1 written at level 3, and so on; PI = parental involvement.

The instructions state: "Please read the descriptions of different parents. Decide how much, in general, each of them is similar to, or different from, your typical situation. Choose one answer." The response options minimize negative wording in these comparisons, for example, "Evaluate your involvement in your child's education compared with X" (X = name of a scenario character). The responses are scored from 1 to 5 with 5 being the highest and hardest-to-attain level of PI: *I am involved in the child's education much more than X = 5; I am involved in the child's education more than X = 4; I am involved in the child's education equally with X = 3; X is involved in the child's education more than I = 2* (Instead of *I am involved less than X* which could be interpreted negatively and might elicit an unintended socially desirable response); and *X is involved in the child's education much more than I = 1*. The scoring values are not seen by the parent. Parent scale scores range from 9 (the lowest level of PI) to 45 (the highest level of PI), excluding the training scenario.

Step 5: *Discussions of the scenarios with psychologists and education experts; cognitive laboratories; subsequent adjustments.* One objective of the cognitive lab think-aloud sessions was to assess how the scenario items were interpreted by parents of primary school students. For example, one parent said, "I can say that I am involved equally with Tatiana. She communicates with teachers and other parents, and sometimes she helps with homework. She is very much like me." Another parent said, "Probably, Fedor is involved more than I. I don't have so much free time for all these activities. And I think it's too much, as it can lead to information overload in the child. Yes, he is involved more." Since some parents interpreted our descriptions too literally, we added and underlined "in general" in the instructions.

Furthermore, instead of providing an extensive explanation of the novel scenario item format in the instruction paragraph, we included a training scenario, which was shorter than the others and designed as an explicitly worded description of a very uninvolved parent. During cognitive labs, people frequently reacted to the training scenario in a similar way: "This is a terrible parent. I am not sure such parents exist. I am definitely involved much more than Ivan." Because it was easy for parents to rate themselves, in comparison, to a very explicitly uninvolved parent, parents felt more confident about how to react to the operational scenarios.

Step 6: *Piloting, analysis, and subsequent adjustments.* The scenarios were piloted on a sample of 388 Moscow parents in autumn 2017. Two scenarios produced confusing responses

Table 3. Scenario Items for the Three Levels of PI.

[Training item] Ivan seldom asks his child about school things and never helps with homework. He never visits school parent–teacher meetings and events.

[High PI] Fedor daily checks how his child copes with homework and whether the child needs his help. Each week he takes his child to a museum or educational tour to illustrate school subjects in real life. Fedor is an active member of a parent committee where he organizes extracurricular activities. Fedor respects his child's opinion about what and how the child wants to learn.

[High PI] Olga every day deals with her child's learning and instead of regular TV in her home they watch educational channels. On weekends they often go on educational sight-seeing tours. Olga attends all school meetings and actively participates in organizing school events. She is serious about her child's opinions on learning goals and is focused on her child's well-being, not just on her child's achievements.

[High PI] Irina decided to drive her child to a distant school which is better than the local one. She creates rich educational environments for her child at home: they very often discuss books, play educational games, go to workshops, theaters, and exhibitions. Irina spends much time on her school parent committee responsibilities. She insists that interest and well-being of her child is more important than the child's grades.

[Medium PI] Anton tries to always have at home books, artist sets, educational games, and sometimes he engages his child in educational discussions. About three to four times a year, Anton takes his child to a museum, planetarium, or zoo. He visits some of the school events and meetings. He insists that his child should achieve certain levels because Anton doesn't want to regret wasting his child's time and opportunities.

[Medium PI] Larisa sometimes finds on the internet or invents for herself interesting ways to explain school subjects to her child. From time to time she takes her child to the library and enrolls in summer educational camps. She attends—one to two school events per year. Larisa expects her child to study and to do not worse than average.

[Medium PI] Tatiana usually monitors how her child is prepared for a new school day, and she helps her gather a school bag. At least once a month she brings her child to a museum, library or theater. She attends the key parent–teacher conferences (first and last in the year) and some school events. Tatiana insists that her child gets good results on some principle subjects (for Tatiana they are English and math).

[Low PI] Maria shows her child what a real adults' responsibilities are such as going to stores, cooking, and paying bills. Maria believes that the current school workload is large enough and parents don't have to add extra educational burdens. Maria almost always skips school events which are often scheduled at inconvenient times. Maria prefers not to interfere with the teacher's educational strategy.

[Low PI] Lidia sometimes asks the child if homework is done. Sometimes she dreams about hiring a special person to look after school things for her child instead of herself. She believes that her life situation does not allow her to participate in school life apart from parenting chats. Lidia thinks that her child should strictly follow the teacher's requirements.

[Low PI] Sergey returns home late so is not involved in the school affairs of his child. On weekends, he and his child rest. Sergey cannot attend school events and parent–teacher conferences. Sergey thinks that school is just one of many life stages, so his child will cope with the school curriculum as well as most other students.

Note. PI = parental involvement.

because they were subtly double-barreled. The parental activity descriptions were then simplified and rephrased.

Step 7: *Final scale administration.* The final PI scenario scale (Parental Involvement SCenario scale [PISC-9]) consists of one training scenario and nine operational scenario items. The scale is presented in the following order: training scenario–medium–high–medium–low–high–low–medium–low–high.

The scale was administered in the Republic of Tatarstan in the autumn of 2017. A total of 1,930 parents of third-graders who participated in the iPIPS longitudinal study responded. This

represents a 98% response rate from the initial iPIPS sample, a remarkable result achieved via constant reminders to parents from schools to fill in the online questionnaires. The sample was representative of the Tatarstan regions, school types, and school sizes. Responding to the PISC-9 takes about 30 to 45 s for each scenario since the cognitive task is more demanding than a typical short-stemmed Likert-type item.

Data Analysis

Rating Scale Model

The data were analyzed under the Rasch rating scale model (RSM; Wright & Masters, 1982) using Winsteps software (Linacre, 2019). The RSM was developed for polytomous items with the expectation that the ordered response options would be interpreted similarly for all items. The model takes the form in Equation 1:

$$\pi_{nix} = \frac{e^{\sum_{j=0}^{x_{ni}}\left[\beta_n - (\delta_i + \tau_j)\right]}}{\sum_{k=0}^{m} e^{\sum_{j=0}^{k}\left[\beta_n - (\delta_i + \tau_j)\right]}},$$

(1)

where π_{nix} is the probability of parent n responding in category x to scenario i (these probabilities generate the expected response of a parent to a scenario); δ_i is the location (scenario difficulty) of scenario i on the PI construct; τ_j is the location (threshold parameter) of the kth transition from one response category to the next for the $m + 1$ rating categories; and β_n is the parameter for a parent's level of parental involvement. Highly involved parents and harder-to-achieve PI scenarios will have positive logit estimates (Ludlow & Haley, 1995). Minimally involved parents and easier-to-achieve PI scenarios will have negative estimates.

Item Analysis

The PI scale's "variable map," presented in Figure 1, demonstrates the relative positions of the scenarios and the parents on the same logit scale. Parents are located to the left of the vertical line, scenarios to the right. Lower involved parents and easier scenarios are at the bottom of the map, higher involved parents and difficult scenarios are at the top. The three left-most columns show the mean scores (the person raw scores divided by 9: a useful quick way of interpreting the score ranges), the raw scores (corresponding to the sum of a parent's answers), and logit values (useful for performing cross-sample or cross-country measurement invariance analyses). The letter M shows the mean of logit measures separately for scenarios and parents; letters S and T show, correspondingly, one and two standard deviations from the mean.

   Rasch measurement scales typically seek a wide spread in the distribution of: (a) the item locations (i.e., difficulty estimates) to cover the full theoretical range of the construct's continuum and (b) the respondent locations (i.e., ability estimates) to capture the full range of human potential on the construct. The standard deviations of those sets of estimates serve as an indication of their spread. Likewise, those person- and item-level estimates have standard errors that capture their variability. Hence, the ratio of a standard deviation (for either the person or the item estimates) to the root mean square standard error of those estimates yields a "separation" index (Wright & Masters, 1982). A separation value of at least two indicates there are at least three levels of meaningful separation along the continuum. The separation in the parents' PI estimates is (2.37); the separation in scenario difficulty estimates is (46.0).
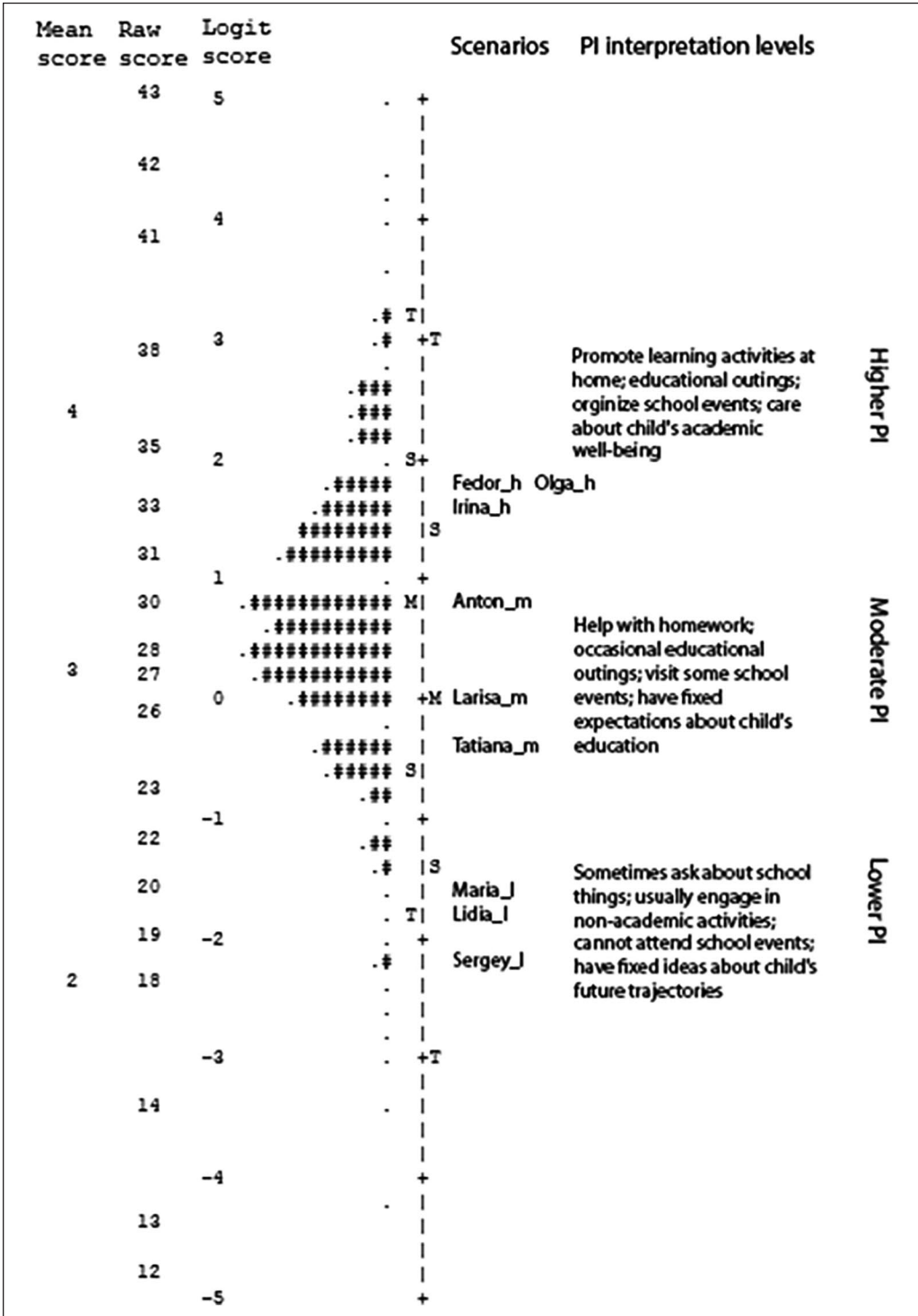
Figure 1. Parent–scenario variable map with related raw scores and mean scores for three levels of PI (high, medium, and low).
Note. Each "#" is 15 parents. Each "." is 1 to 14 parents. PI = parental involvement.

The purpose of a variable map is twofold: (a) to show what it means to progress from lower levels of a variable (the PI construct) to the upper levels and (b) to provide a confirmatory test of the instrument developer's hypothesized structure of the construct. As seen in Figure 1, scenarios intended to describe higher involved parents (marked with letter "h") do have the higher difficulty estimates; scenarios intended to describe lower involved parents ("l") have lower estimates, and scenarios which were planned to describe parents between them, have medium difficulty estimates ("m"). Illustrating a critical feature of Rasch measurement, the scenarios successfully capture the *a priori* hypothesized continuum of parenting behaviors ranging from those representing lower levels of involvement to those representing higher levels of involvement.

Based on the map's scenario locations, and the apparent clustering of the scenarios into three relatively distinct levels of parental involvement, it is possible to provide rich meaningful interpretations of the parents' scores. The lower level of involvement (parents with raw scores ranging from 9 to 22) is associated with "easier to score high on" scenarios. Parents in this section of the PI continuum spend time with their children in daily routines and recreation like watching TV and going to a local playground. Their life situation prevents them from deeper engagement with school affairs and they often form their attitudes about their children's school results using social norms ("Parents should earn, teachers should teach"), and their own past experience ("My parents did not attend school events and I grew up well enough").

At the medium level of involvement (scores ranging from 23 to 31) parents demonstrate a special interest in the academic life of their children. For example, they regularly monitor homework and sometimes tutor their children in difficult subject topics. Several times a year they organize cultural or educational outings to museums and workshops. In general, they have strong opinions about education which they stick to ("School success is important," "Mathematics is a must"), and they work with their children to support these beliefs.

Highly involved parents (those who score 32 or more) are very concerned about and engaged with their children's academic development. They monitor homework every day and teach their children how to organize themselves. They intensively tutor their children and do activities outside of the required curriculum (e.g., teach their child coding, learn with the child a new poem each week). They frequently organize cultural or educational outings, enroll their children in educational courses and use every possibility to relate academic knowledge to real life. These parents often take on responsibility as the head of parental committees to help organize school events, and they intentionally build communications with the teacher. They pay considerable attention to the well-being of their children and their interests; and moderate their child's educational life accordingly.

Fit analysis—Scenarios. In our goodness-of-fit analysis for both parents and scenarios, we employed unstandardized, weighted mean square statistics (the "Infit MNSQ" in WINSTEPS) and unweighted mean square statistics (the "Outfit MNSQ") (Wright & Masters, 1982). These statistics are based on the residuals produced by the differences between the observed responses provided by the parents on each scenario and their expected responses under the Rasch rating scale model (Ludlow, 1983; Wright & Masters, 1982). Their expected values are 1.0. Outfit highlights particularly unusual "outlier" responses to as few as a single scenario (or parent); infit highlights relatively consistent inconsistent response patterns across the full set of scenarios (or parents).

We used a liberal criterion of +1.3 to flag potential "misfit" problems (rather than the conventional criterion of 1.5; Wright & Linacre, 1994) because this relatively low criterion will minimize missing a potential problem (i.e., a protection against Type II error)—a crucial point in the development of a new instrument. Not only do these statistics highlight instances of unexpected responses but they also serve as tests of the equal discrimination assumption. Items for which high and low scoring people have responded as expected will have fit statistic values varying around 1.0. Items for which the high and low scoring people have responded in patterns opposite

Table 4. Item Difficulty Estimates and Fit-Statistics.

Scenario	Measure	SE	Infit MNSQ	Outfit MNSQ
Fedor_h	1.78	0.04	1.05	1.05
Olga_h	1.75	0.04	0.96	0.95
Irina_h	1.56	0.04	1.09	1.1
Anton_m	0.75	0.03	0.76	0.77
Larisa_m	−0.02	0.03	0.88	0.91
Tatiana_m	−0.36	0.03	0.94	0.99
Maria_l	−1.51	0.03	1.06	1.14
Lidia_l	−1.75	0.04	1.11	1.19
Sergey_l	−2.2	0.04	1.03	1.17

Note. MNSQ = mean square.

to what was expected (e.g., some high scoring people unexpectedly scored low on an easy scenario, or some low scoring people unexpectedly scored high on a hard scenario) will have high positive fit statistics. High positive fit statistics are associated with poor discrimination and, consequently, have poor item-total correlations (Wright & Masters, 1982). All scenarios demonstrated good fit statistics (see Table 4) according to the conventional criteria of 0.7 to 1.5 (Wright & Linacre, 1994).

Response category functioning. The category characteristic curves (Figure 2) display the probability of responding in each scoring category based on the difference between any parent's PI estimate and any scenario's difficulty estimate (Wright & Masters, 1982). The patterns for these curves indicate that the item categories functioned as intended: (a) category difficulty threshold estimates are well distinguished (−4.6, −0.71, 1.51, 3.8); (b) mean PI measures of those who responded in each category increased gradually from the first scoring category to the fifth (−3.1, −1.48, 0.38, 2.42, 3.9); and (c) all response category fit statistics were within the 0.7 to 1.5 range (0.86–1.25) (Linacre, 2002).

Dimensionality analysis. A principal component analysis of the Rasch model residuals checked the unidimensionality assumption of Rasch measurement (Ludlow, 1983; Rasch, 1960/1980). A principal component analysis rather than a common factor analysis is performed because the residual matrix is assumed to consist of only error variance. The plot of the first two residual components was consistent with random data (i.e., the scenarios were distributed in a circular pattern around the origin) although the first component eigenvalue was slightly higher than the 1.0 expected under a parallel analysis of random data (O'Connor, 2000). Subsequent analysis of the first component residual variation revealed that the two more difficult scenarios were slightly correlated ($r = .02$) while the two least difficult scenarios also correlated ($r = .04$; neither of which was statistically significant).

Theoretically, highly involved parents are different from lower involved parents in terms of behavior, financial situation, and cultural capital. In the process of instrument development, we paid special attention to avoid socioeconomic status (SES) indicators in the wording of scenarios such as avoiding emphasizing financial investments of parents in tutoring, purchasing books or educational games, and other income-related activities. However, it was impossible to completely avoid SES issues, because free time and knowledge that people can invest in their children also relates to their educational level and work schedule. This modest residual clustering of the higher and lower pairs of scenarios was accepted as tolerable. In addition, the global log likelihood chi-square test of the data fit to the Rasch rating scale model was nonsignificant for, $\chi^2(26, 629) = 26,495.4, p = .72$.
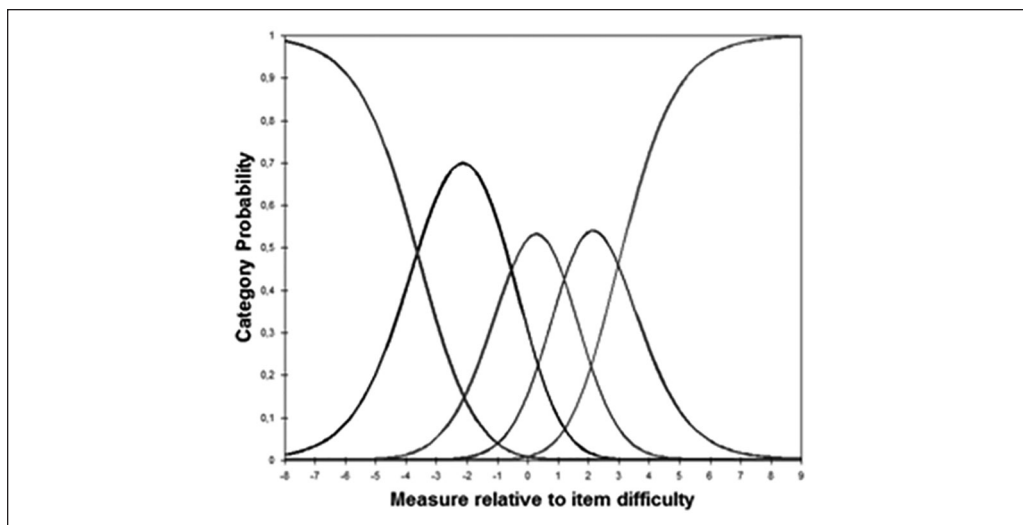
Figure 2. Category characteristic curves.

Fit analysis—Parents. We also investigated misfitting parents who gave unexpected responses based on their level of PI (i.e., some overall low scoring parents scored high on a relatively hard scenario and some overall high scoring parents scored low on a relatively easy scenario). For example, one particularly high scoring parent with an expected response of 5 on "Maria2_l" responded to this scenario with a 1—a very unexpected response that triggered a misfit alert. This single instance of an unexpected response produced a large standardized residual (−9.0) which when squared produced an outfit statistic of (9.9). In other cases, parents had expected responses of four but responded with three, and other combinations of relatively slight differences occurred too. Based on our low criterion for misfit (infit or outfit >1.3), profiles of about 10% of the parents showed they gave varying degrees of inconsistent responses like these under the RSM expectations. Only 6% of the sample exceeded the traditional criterion of 1.5 (Wright & Linacre, 1994).

Because our measurement intention was to build a scale with data that fit the Rasch model as opposed to building an IRT model that would best recover the response data, we initiated an investigation of the possible sources of parent misfit. Hence, we compared the misfit subsample (defined as infit >1.3, $n = 193$) with the rest of the sample on several demographic variables: (a) family financial situation (a subjective opinion of the parent containing six categories from "We live very cost-conscious (economically) and sometimes struggle to have money for food" to " We have no financial problem and can buy a car or house without getting a credit or mortgage"; (b) language they speak at home (Russian or other); (c) language they use to discuss school-related things with their child (Russian or other); (d) mother's education (at least college degree vs. no college degree); (e) the level of education they expect their child will achieve in the future (at least college degree vs. no college degree); and (f) number of books at home ("five or less" or "26 or more").

Chi-square tests of independence (all cells had expected frequencies greater than five) showed no significant differences based on financial situation, and borderline to significant differences for language at home ($p = .06$: misfitting parents chose other than Russian languages more often). Differences on other variables were, however, significant. Parents from the misfit subsample responded significantly more often that they discussed school-related things with their children in other than Russian languages ($p < .05$), and had less than 25 books at home ($p < .01$).

Mothers from the misfit subsample less often reported that they had a college degree ($p <. 01$) or expected their child to obtain a college degree ($p < .01$). Interestingly, standardized reading and math achievement test scores of the children in the two subsamples did not differ significantly (t-test, $p > .05$).

To test the properties of the PISC-9 based on parents for whom the scale was most appropriate, the subsample of misfitting parents was deleted and the analyses were rerun. This step is consistent with an influential observation analysis performed in traditional statistical applications (Belsley et al., 1980) and the practice of trimming in economic studies (Winer, 1971).

Results from the trimmed sample analysis include: the variable maps for the original and trimmed data were essentially identical in their scenario location estimates but the standard errors for the trimmed estimates were smaller; the parent separation index improved from 2.14 to 2.37; the scenario separation index improved from 41.3 to 46; the category response thresholds reported earlier widened from the full sample estimates of ($-3.6, -0.54, 1.2, 2.98$) and the mean parent PI estimates within the scoring categories increased from ($-2.3, -1.1, 0.35, 1.9, 3.0$); the percent of misfitting people in the trimmed sample was 4% versus 10% for the full sample; the global log likelihood chi-square test of the data-to-model fit improved from, $\chi^2(32, 942) = 32, 806.9$, $p = .70$; and the Cronbach's alpha for the PISC-9 scale improved from .82 to .85. The results reported in this study are based on this trimmed sample of 1,737 parents.

Differential item functioning analysis. Differential item functioning (DIF) was checked using the Mantel–Haenzel method (Zwick et al., 1999) on several variables: (a) financial situation (0: lower, 1: higher); (b) educational expectations for the child (1: parents expect their child to achieve a college degree, 0: parents do not expect their child to achieve a college degree); (c) mother education (0: no college degree, 1: with college degree); and (d) language at home (1: Russian, 0: other). Based on the absolute DIF contrasts not exceeding 0.43 (Linacre, 2017), no DIF was evident.

Validating the scenario scale against traditional scales. To compare the PISC-9 scale scores against traditional PI scale scores, we added into the parental questionnaire several Likert-type items on school-based and home-based PI (see Table 5). One out of four school-based PI items and four out of nine home-based PI items were kept from the adaptation of the APQ (Frick et al., 1999) and the rest were the original items that showed satisfactory characteristics on the previous wave of the iPIPS project (Antipkina, Lyubitskaya & Nisskaya, 2018).

The Cronbach's alpha for the home-based scale is .62 and the school-based alpha is .72. Although these alphas are relatively low, the scales are intended only for discriminate validity purposes (if the number of items within each scale are doubled, for example, the Spearman–Brown prophecy formula [Crocker & Algina, 1986] estimates alphas of .77 and .84, respectively). The correlations between PISC-9 scores and the home-based and school-based scores are .36 and .499, respectively (see Table 6). Campbell and Fiske (1959) state that two measured concepts are likely to exhibit discriminant validity if the disattenuated correlations are less than .85. Here, the respective discriminate validity coefficients are .51 and .65.

Functioning of a shortened version of the scenario scale. In a separate, small-scale study, we distributed a shortened version of the PI scenario scale to a sample of 490 parents of fourth graders in Krasnoyarsk, Russia. One scenario was removed from each of the three levels of the PI scale to reduce content redundancy and response burden. The shortened scale includes six items (two higher [Olga_h, Fedor_h], two medium [Larisa_m, Tatiana_m], and two lower level [Sergey_l, Maria_l] scenarios) and the training item. The shortened scale (PISC-6) takes about 2 min less than the full PISC-9.

Table 5. Traditional PI Items Used in Comparison With Scenarios.

School-based PI

> You ask your child's teacher or share with her some information
> You bring things to school
> You participate in organizing school events
> You volunteer to help with special activities that your child is involved in

Home-based PI

> You help your child with subjects that are difficult for her or him
> You help your child prepare for the next day
> You do with your child something interesting for her or him
> You take your child to the library
> You take your child to museums, exhibitions, and so on
> You take your child to extracurricular activities
> You ask your child what his or her plans are for the coming day
> You help your child with his or her homework
> You ask your child about his or her day in school

Note. Scoring categories: "*never*"=1; "*1 or 2 times a year*"=2; "*almost every month*"=3; "*almost every week*"=4; "*More than once a week*"=5. PI = parental involvement.

We replicated the previous analyses and the results show that the shortened scenario scale is unidimensional (nearly identical empirical and parallel analysis simulated residual eigenvalues); has excellent scenario fit-statistics (infit and outfit: [0.97, 0.96]; [0.93, 0.92]; [0.88, 0.90]; [1.1, 1.1]; [0.92, 0.91]; and [1.1, 2.0], respectively); scenarios are distributed along the continuum in their expected order; the Cronbach's alpha is .79; and the global log likelihood chi-square test of the data-to-model fit is nonsignificant for, $\chi^2(4, 508) = 4, 482.3$, $p = .56$. Using the larger trimmed sample, the correlation between the pairs of parent logit PI estimates on the nine-scenario and six-scenario versions is .973 ($p < .005$).

Figure 3 presents the PISC-6 variable map. The scenarios are well spread in their definition of the PI construct—a clear ladder-like progression is evident from the easiest scenario (Maria_l) up to the hardest one (Olga_h). There are only a few extremely high or low PI scores—a fact that suggests parents took the task seriously and did not exaggerate their involvement one way or the other. Although the content of all nine scenarios in Figure 1 (PISC-9) produces a richer description of the types of activities and behaviors associated with different levels of parental involvement than does PISC-6 in Figure 3, PISC-6 is quick and psychometrically sound.

Discussion

This article describes the development of the PISC-9. The development of this scale was prompted by concerns that the construct of parental involvement allows for multiple definitions and operationalizations and is often too narrowly focused on specific aspects of parental involvement.

The objective of our scenario-based PI scale is to measure PI more holistically, instead of following a typical approach of selecting items that maximize Cronbach's alpha. In contrast to such an approach, we have shown that a scale based on Rasch measurement principles and Guttman facet theory design can measure a range of authentic PI behaviors. Unlike a more traditional scale, where a score is often just treated as a predictor or outcome in a correlation or regression study with little if any substantive value (other than, perhaps, as a percentile or T-score), here, we show how a score may be interpreted in terms of the most likely types of involvement a parent is
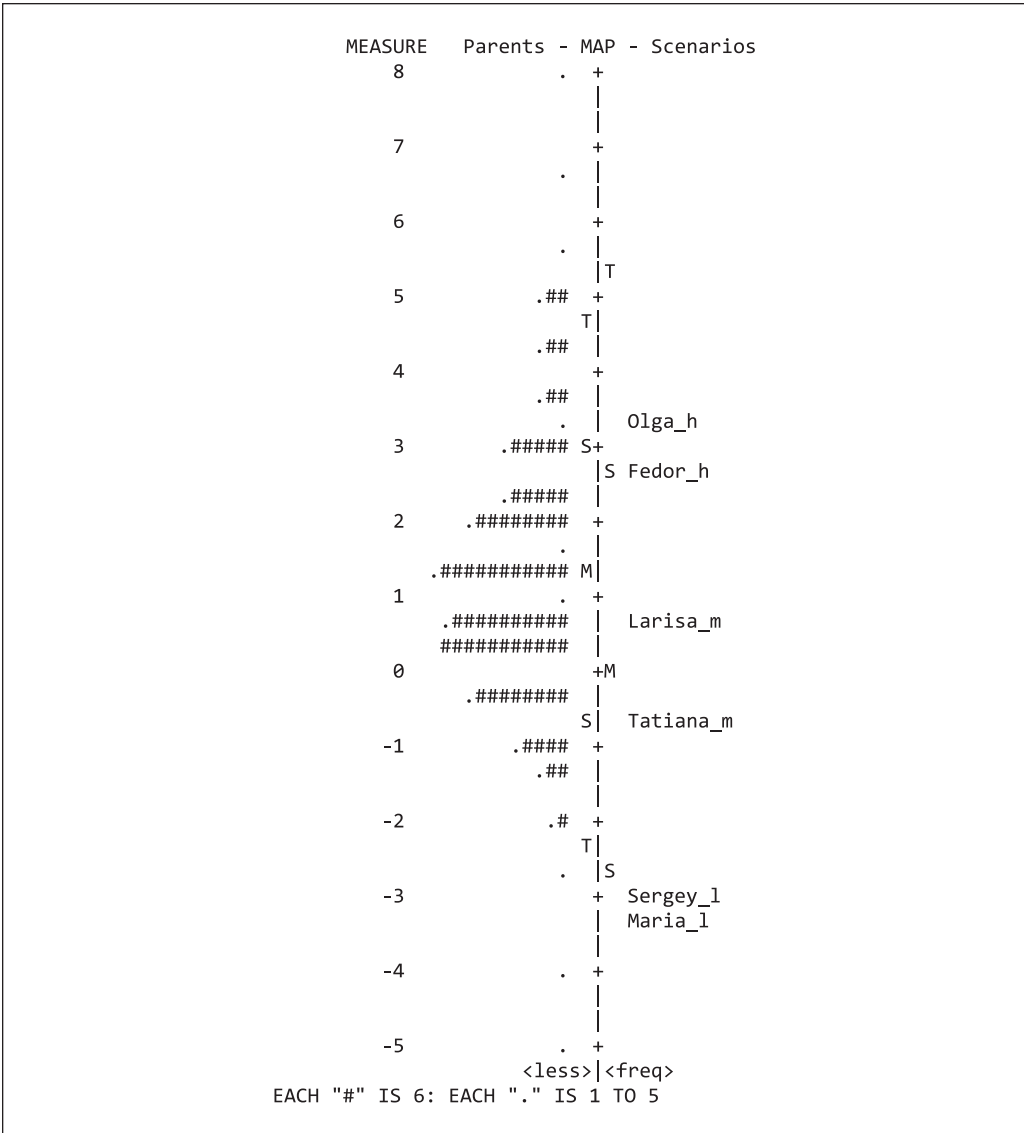
```
            MEASURE    Parents - MAP - Scenarios
               8               .  +
                                  |
                                  |
               7                  +
                               .  |
                                  |
               6                  +
                               .  |
                                  |T
               5             .##  +
                                 T|
                             .##  |
               4                  +
                             .##  |
                               .  |   Olga_h
               3           .##### S+
                                  |S Fedor_h
                            .##### |
               2          .####### +
                               .  |
                  .########### M|
               1               .  +
                    .########## |   Larisa_m
                    ########### |
               0                  +M
                     .######## |
                                 S|   Tatiana_m
              -1           .#### +
                            .## |
                                 |
              -2             .# +
                                T|
                              .  |S
              -3                  +   Sergey_l
                                  |   Maria_l
                                  |
              -4             .  +
                                  |
                                  |
              -5             .  +
                            <less>|<freq>
            EACH "#" IS 6: EACH "." IS 1 TO 5
```

Figure 3. Parent–scenario variable map for the shortened PISC-6 scale.
Note. Each "#" is six parents. Each "." is one to five parents.

engaged in. This opportunity to provide a meaningful interpretation of a score offers not only diagnostic information about location status on the PI construct but also the opportunity to substantively measure change. In the present situation, this is equivalent to asking "what does it take for a parent to become more involved?"

It is important to note, that we did not try to define "the best" or "normal" way to be involved in a child's education. Each scenario contains a combination of typical aspects (facets) of parental involvement chosen according to theoretical and empirical reasons. Thus, the scenario format provides a rich lived-experience interpretation which allows us to describe many varieties of parental behavior along a potentially infinite continuum of parental involvement.

These interpretational advantages are supported by research comparing scenario scores to traditional scale scores measuring comparable constructs. In a comparison of Productive Engagement

Table 6. Correlations Between PI Scenario Scale and Two Traditional Scales of Home-Based and School-Based PI.

Scales	School-based PI (Likert-type)	Home-based PI (Likert-type)	Scenarios PI
School-based PI (Likert-type)	I	0.592	0.360
Home-based PI (Likert-type)		I	0.499
Scenarios PI			I

Note. PI = parental involvement.

Portfolio (PEP) scenario scores versus Utrecht Work Engagement Scale (UWES) Likert-type scores (Schaufeli et al., 2006), the PEP scores did not reflect the ceiling and negatively skewed effects seen in the UWES scores (Ludlow, Matz-Costa & Klein, 2019). Furthermore, a comparison of PEP scenario scores versus Likert-type and semantic differential scores created from re-expressions of the PEP scenarios demonstrated richer, more detailed interpretations of productive work engagement levels through the scenario descriptions (Matz-Costa et al., 2014).

We also have shown that (a) an explicit training item can successfully introduce this new format to respondents without long explanations in instructions and (b) a shortened scenario scale (six items plus one training item) can function very well. The robustness of the PISC-6 scale will be useful when long parent questionnaires consist of several scales and a short PI scale is desired as a way to minimize construct irrelevant variance introduced through respondent fatigue or lose of motivation.

Depending on the circumstances either the PISC-9 or PISC-6 may be used as an quantitative index of PI. The PISC-9, for example, is presently included in the Russian iPIPS large-scale assessment. Other audiences may also benefit from these scales. For example, parent–teacher meetings can use the scenarios as guidelines for opening conversations about what parents can do at home to become more involved; and teacher professional development sessions can likewise use the scenarios as ways of considering how to involve parents in school-based activities.

Limitations of the Study

The scenario format may be more demanding of the reading and language skills of respondents than traditional short-stemmed Likert-type items. This could cause respondent confusion and misinterpretations. It is also possible that male and female scenario names trigger gender stereotypes in some respondents, which may, interfere with their answers. Both situations would introduce construct irrelevant variance into the assessment.

Further Directions of the Study

It would be informative to check the invariance of the scale's structure, and its operationalization of the PI construct, in other countries/cultures. Despite the fact that the scenario format is more demanding to employ than traditional item and scale development procedures, it allows for a flexible definition of the construct. This means that context-specific descriptions of parents and the things they do regarding their child's education may be adapted detail-by-detail, depending on the cultural situation, through systematic manipulation of the facets and their levels in the sentence mapping framework.

In addition, potential gender stereotypes should be studied further along with, among aspects of scenario development: how many scenarios are the minimum for a well-defined continuum? At what points are there too many facets and sentences in a scenario? Must the facets stay in the same order of presentation for all scenarios?

Finally, what are the educational benefits of the PISC-9 and PISC-6 scales for the children? Or, how does the measurement of parental involvement translate into actions that benefit the educational opportunities of the child? These questions suggest a rich research future for these scales, in particular, and this measurement approach, in general.

Declaration of Conflicting Interests

Funding

ORCID iD

Inna Antipkina https://orcid.org/0000-0003-4865-3433

Note

1. iPIPS procedures were approved by the ethical committee of the Higher School of Economics University. The procedures for collecting, storing, and analyzing data follow the requirements of Russian legislation on the protection of personal information (Law "On Information, information technologies, and protection of information from 27.07.2006 No 149-FZ").

References

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social and health sciences*. Springer.

Antipkina, I., Lyubitskaya, K., & Nisskaya, A. (2018). Parental Involvement in Schools. *Voprosy obrazovaniya, 4* (eng), 230–260.

Bakker, J. T. A., & Denessen, E. J. P. G. (2007). The concept of parent involvement: Some theoretical and empirical considerations. *International Journal About Parents in Education*, *1*, 188–199.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley.

Borg, I., & Shye, S. (1995). *Advanced quantitative techniques in the social sciences: Vol. 5. Facet theory: Form and content*. SAGE.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Chang, W. C., Ludlow, L. H., Grudnoff, A., Ell, F., Haigh, M., Hill, M. & Cochran-Smith, M. (2019). Measuring the complexity of teaching practice for equity: Development of a scenario-format scale. *Teaching and Teacher Education*, *82*, 69–85.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

Engelhard, G. (2012). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.

Fantuzzo, J., Tighe, E., & Childs, S. (2000). Family Involvement Questionnaire: A multivariate assessment of family participation in early childhood education. *Journal of Educational Psychology*, *92*, 367–376.

Frick, P. J., Christian, R. C., & Wootton, J. M. (1999). Age trends in the association between parenting practices and conduct problems. *Behavior Modification*, *23*, 106–128.

Garbacz, S. A., & Sheridan, S. M. (2011). A multidimensional examination of New Zealand family involvement in education. *School Psychology International*, *32*, 600–615.

Grolnick, W. S., Benjet, C., Kurowski, C. O., & Apostoleris, N. H. (1997). Predictors of parent involvement in children's schooling. *Journal of Educational Psychology*, *89*(3), 538–548.

Guttman, R., & Greenbaum, C. W. (1998). Facet theory: Its development and current status. *European Psychologist*, *3*(1), 13–36.

Hornby, G. (2005). *Improving parental involvement*. Continuum.

Jeynes, W. (2010). *Parental involvement and academic success*. Routledge.

Kardanova, E., Ivanova, A., Sergomanov, P., Kanonire, T., Antipkina, I., & Kayky, D. (2018). "Patterns of first-graders" development at the start of schooling: Cluster approach based on the results of iPIPS project. *Educational Studies*, *1*, 8–37.

Kastberg, D., Roey, S., Ferraro, D., Lemanski, N., & Erberber, E. (2013). *US TIMSS and PIRLS 2011 technical report and user's guide* (NCES 2013-046). National Center for Education Statistics.

Lareau, A. (2011). *Unequal childhoods: Class, race, and family life*. University of California Press.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*(1), 85–106.

Linacre, J. M. (2017). *Winsteps* (Version 4.0. 0) [Computer software]. Winsteps.com.

Linacre, J. M. (2019). *Winsteps® Rasch measurement computer program*. Winsteps.com.

Ludlow, L. H. (1983). *The analysis of Rasch model residuals* (Doctoral dissertation). University of Chicago, Chicago IL.

Ludlow, L. H., Braun, H., O'Keefe, T., Howell, B., Anghel, E., & Szenday, O. (2019). *The Boston College Living a Life of Meaning and Purpose (BC-LAMP) Portfolio: A series of Rasch/Guttman scenario scales*. Paper presented at the American Educational Research Association annual meeting. Toronto, Canada.

Ludlow, L. H., & Haley, S. M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement*, *55*, 967–975.

Ludlow, L. H., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., & James, J. B. Author. (2014). Measuring engagement in later life activities: Rasch-based scenario scales for work, caregiving, informal helping, and volunteering. *Measurement and Evaluation in Counseling and Development*, *47*(2), 127–149.

Ludlow, L. H., Matz-Costa, C., & Klein, K. (2019). Enhancement and validation of the Productive Engagement Portfolio-Scenario (PEP-S8) Scales. *Measurement and Evaluation in Counseling and Development*, *52*(1), 15–37, https://doi.org/10.1080/07481756.2018.1497430

Ludlow L. H., Reynolds K., Baez-Cruz M., & Chang, W.-C. (in press). Enhancing the interpretation of scores through Rasch-based scenario-style items. In (Harbaugh AG: Ed). *Basic Elements of Survey Research in Education: Addressing the Problems Your Advisor Never Told You About*.

Manz, P. H., Fantuzzo, J. W., & Power, T. J. (2004). Multidimensional assessment of family involvement among urban elementary students. *Journal of School Psychology*, *42*(6), 461–475.

Matz-Costa, C., James, J. B., Ludlow, L., Brown, M., Besen, E., & Johnson, C. (2014). The meaning and measurement of productive engagement in later life. *Social Indicators Research*, *118*(3), 1293–1314. doi: 10.1007/s11205-013-0469-6

McNeal, R. B., Jr. (2014). Parent involvement, academic achievement and the role of student attitudes and behaviors as mediators. *Universal Journal of Educational Research*, *2*(8), 564–576.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396–402.

Pomerantz, E. M., Moorman, E. A., & Litwack, S. D. (2007). The how, whom, and why of parents' involvement in children's academic lives: More is not always better. *Review of Educational Research*, *77*(3), 373–410.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* [Copenhagen: Danish Institute for Educational Research, expanded edition with foreword and afterword by B. D. Wright]. The University of Chicago Press. (Original work published 1960)

Reynolds, K. (2019, April). *A Rasch/Guttman-based scenario scale for college faculty out-of-class availability* [Paper presentation]. American Educational Research Association Annual Meeting, Toronto, ON, Canada.

Schaufeli, W. B., Bakker, A. B., & Salanova, A. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, *66*, 701–716.

Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, *73*(2), 445–460.

Shen, T., & Ludlow, L. H. (2018, April 16). *The community readiness and participation scenario scale: Unidimensional or multidimensional?* Paper presented at the American Educational Research Association annual meeting. New York, NY.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*(4), 529–554.

Walker, J. M., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental involvement: Model revision through scale development. *The Elementary School Journal*, *106*(2), 85–104.

Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review*, *66*(3), 377–397.

Wilson, M. (2004). *Constructing measures: An item response modeling approach*. Routledge.

Winer, B. J. (1971). *Statistical principles in experimental design*. McGraw-Hill.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370–371.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Mesa Press.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*, *36*, 1–28.