Applying beliefs and resources frameworks to the psychometric analyses of an epistemology survey

Sevda Yerdelen-Damar,[1] Andrew Elby,[2] and Ali Eryilmaz[3]

[1]*Secondary Science and Mathematics Education, Yüzüncü Yıl University, 65080, Van, Turkey*
[2]*Department of Physics and Department of Curriculum and Instruction, University of Maryland, College Park, Maryland 20742, USA*
[3]*Secondary Science and Mathematics Education, Middle East Technical University, 06800, Ankara, Turkey*

This study explored how researchers' views about the form of students' epistemologies influence how the researchers develop and refine surveys and how they interpret survey results. After running standard statistical analyses on 505 physics students' responses to the Turkish version of the Maryland Physics Expectations-II survey, probing students' epistemologies and expectations, we interpreted the results through two different theoretical lenses, the beliefs perspective and the resources perspective. We showed that the beliefs and resources frameworks provided different interpretations of the psychometric analyses, leading to different conclusions about how the survey results should be interpreted and how the survey should be improved.

I. INTRODUCTION

Students' personal epistemologies, their understandings of the nature of knowledge, knowing, and learning [1–3], influence their success at learning [4–9] and their approaches to learning [4,10–13] in science and mathematics. Although many researchers agree upon the importance of students' epistemologies to their learning, there is some controversy on the form of epistemologies in students' minds. In the personal epistemology literature, most researchers conceptualize students' personal epistemologies as made up of relatively coherent and stable cognitive structures [14], such as developmental stages [15,16], intuitive theories [2,17], or systems of quasi-independent beliefs along multiple dimensions [18]. However, some recent studies question this assumption and propose the idea that personal epistemologies consist of finer-grained cognitive elements—epistemological resources—whose patterns of activation depend on context [11,14,19]. Researchers' views about the cognitive form of students' personal epistemologies are important because they affect how data are collected and interpreted. For example, a researcher who conceives of epistemologies as intuitive theories can ask interview subjects fairly decontextualized questions about their views of physics; contextualization is not needed to unearth a subject's theory. Other researchers, in contrast, might suspect that differently contextualized questions would elicit different responses and would plan their interviews accordingly.

In this study, we discuss how researchers' (perhaps tacit) views about the form of students' epistemologies influence how they develop and refine surveys and how they interpret results. After running standard statistical analyses on physics students' responses to a widely used survey probing epistemologies (and a related construct, expectations), we interpret the results of those analyses through two different theoretical lenses, the beliefs perspective (our shorthand for stages, theories, or beliefs) and the resources perspective. By doing so, we not only establish psychometric properties of an American survey when translated into Turkish, but also help to clarify the differences between those two theoretical perspectives by demonstrating how they affect survey construction and interpretation.

A quick, nonphysics example helps us summarize the argument we are going to make in detail below. Consider these two Likert scale (agree-disagree) items designed to probe people's beliefs about how progressive the American tax system should be.

(1) If a worker works harder and therefore makes more money than she did last year, she should *not* be forced to pay a higher percentage of her income in taxes to the government.

(2) If a hard worker is downsized and therefore earns less money this year than she earned last year, she should pay a lower percentage of her income in taxes.

It is imaginable that many respondents would agree with both of these statements, and let us suppose that is the case. How should this result inform refinement and/or interpretation of these survey items?

If a researcher assumes that people have coherent *beliefs* about whether the tax system should be progressive (i.e., whether higher-income workers should pay a higher percentage of their income in taxes), then the two items

should elicit the same view about whether taxes should be progressive; people who agree with (1) should disagree with (2). So, if many respondents agree with (1) *and* (2), it indicates that one or both survey items is flawed, i.e., one or both items fail to elicit the respondent's real belief about progressive taxation. The researcher would therefore need to refine or replace one or both of those items.

In contrast, if a researcher assumes that people *might* have context-dependent views (as opposed to coherent beliefs) about taxation, widespread agreement with both (1) and (2) can be interpreted as indicating actual context dependence in—rather than flawed measurement of— people's thinking about taxation. For instance, maybe the cues in item (1) about income level being *earned* trigger different ways of thinking about taxation than do the cues in question (2) about income level being subject to forces beyond the worker's control. In that case, the researcher might want the survey to include both (1) and (2), to explore those context dependencies.

By playing out this same argument with (actual) Maryland Physics Expectations-II (MPEX-II) results, we hope to (i) clarify the differences between the *beliefs* and *resources* perspectives and (ii) illustrate the divergent methodological implications of these two theoretical perspectives, for survey construction and interpretation. In general, we discussed that if we approach the results of psychometric analyses of the survey items in the beliefs perspective, the items which are poorly correlated to the other items hypothesized to be assessing the same construct should be eliminated or revised because students' epistemological views are coherent and stable. However, if we approach the results in the resource perspective, those uncorrelated items can be used as evidence that different contextual cues embedded in different survey items can activate different sets of epistemological resources.

A. Review of theoretical frameworks for describing student epistemologies

In this section, we briefly review the beliefs framework and then the resources framework.

1. Beliefs framework

Three distinct cognitive frameworks fall under this perspective: epistemologies as (1) developmental stages, (2) personal theories, or (3) systems of quasi-independent beliefs (that might not form a ''theory''). According to epistemological stage theorists, people's epistemologies proceed in a predetermined order through unidimensional developmental stages [e.g., from absolutist to multiplist (relativist) to a more expert ''evaluativist'' stage in which knowledge is viewed as constructed though knowledge claims are more warranted than others] [15,16]. The personal theories framework conceptualizes students' epistemologies as quite coherent systems of beliefs, but with a much broader space of possible belief systems than the

strict stage theorists would allow (e.g., a student having sophisticated views about the coherence of knowledge but still believing it comes largely from authority and direct observation rather than construction or inference) [2]. Finally, under the quasi-independent beliefs framework, students have a coherent epistemological belief along each of a small number of dimensions, but those beliefs need not cohere into a theory [18]. Although these three frameworks differ in how they conceptualize personal epistemology, they all share the idea that students' personal epistemologies correspond to comparatively context-independent, stable cognitive structures which an individual either does or does not possess [19], analogous to the ''conceptions'' and ''misconceptions'' discussed in much physics education research [20]. Put another way, these three frameworks all hold that students ''have'' epistemological beliefs that are coherent at the level of the individual beliefs (and perhaps at other levels of organization, too). There is a great deal of research advocating this idea [2,17,21,22]. For instance, Schommer and Walker [22] conducted a study to test the assumption that epistemological beliefs are domain independent. They concluded that students tend to have consistent epistemological beliefs across domains, such as social science and mathematics. Moreover, Smith and Wenk [17] interviewed 35 college freshmen to investigate whether there is coherence in students' thinking about epistemological issues using three different types of probes. They argued that their data indicated coherence in students' epistemological thinking.

Though some beliefs advocates have recently argued that students' epistemologies vary across domains [21,23,24], they still consider them as coherent, stable cognitive structures within a particular domain.

2. Resources framework

On the other hand, several studies also challenge the idea that students' epistemologies are coherent stable beliefs across contexts even within a domain [5,9,11,14,25–27]. For example, Leach *et al.* [25] compared students' responses across two kinds of diagnostic questions that probed students' views about the extent to which theory drives the collection and interpretation of experimental data. Answering decontextualized Likert scale items, most displayed a different epistemological stance than they did when commenting upon a concrete example of the same data being interpreted in two different ways. Similarly, Sandoval and Morrison [27] interviewed eight high school students before and after a 4-week technology-supported inquiry unit on natural selection and evolution. They noted that, even within a given interview, individual students' responses to different questions varied widely across epistemological levels.

Hammer and Elby [19] proposed an alternative framework to account for the variability in students' personal epistemologies across contexts. According to this

framework, the "atoms" of students' naive epistemologies are epistemological resources analogous to diSessa's [28] phenomenological primitives (*p*-prims) in intuitive physics.

In this view, students' epistemological resources are less formal than epistemological beliefs, and their activations depend heavily on context. For example, a young child can respond differently to the question "how do you know that?" depending on the situation. When asked how she knows what they are going to have for dinner, she may respond that her mother told her. This answer indicates that the child has intellectual resource(s) for thinking of knowledge as a kind of "stuff" that can be passed from one person to the next. On the other hand, if she is asked how she knows $3 \times 5 = 15$, she might answer that she added $5 + 5 + 5$. This response shows the child has resource(s) for thinking of knowledge as developed from other knowledge. The child's capability of providing epistemologically distinct responses to the same question in different contexts indicates the existence of a variety of different epistemological resources [19]: different contexts activate the child's different epistemological resources for understanding the source of knowledge.

This variation does not mean that personal epistemologies are haphazard and incoherent. On the contrary, the resources perspective accounts for *local coherences* in students' epistemologies [29,30]. According to the resources perspective, a given context can evoke the locally coherent activation of a network of epistemological resources [29,30], leading to belieflike coherence in the student's epistemological stance in that context. "Locally coherent" means the activations of the individual resources are mutually reinforced by each other and/or by features of the context. By this account, other contexts can evoke different epistemological local coherences, i.e., a different "belief." For instance, a student sitting in physics lectures, cued by the professor presenting information and by the other students taking notes and also cued by previous experiences with school science, may activate (consciously or unconsciously) a locally coherent set of epistemological resources that form a "transmissionist" epistemological stance. But the same student, participating in a collaborative small-group learning activity with classmates and scaffolded by curriculum that helps students build their own understandings, might get cued into a more "constructivist" stance.

Now of course, an epistemological "belief" that begins as a (mere) local coherence can, through repeated activation and conscious reflection, become a full-fledged, stable belief. Among experts, such epistemological beliefs (or theories) are undoubtedly common. But among novices, the resources perspective predicts that the "beliefs" displayed in a given situation are often local coherences whose activation and stability depend on the context.

II. METHOD

A. Participants

The study included 505 (270 female, 235 male) tenth grade students from four public high schools in a district of Ankara, the capital of Turkey. Their ages ranged from 15 to 17. The schools share the same physics curriculum and courses. When our data were collected, the physics courses generally relied upon teacher-centered instruction.

B. Instrumentation

The MPEX-II survey is mainly used to probe students' epistemological stance in physics. It was developed by Elby *et al.* [31] from the original Maryland Physics Expectations (MPEX) survey [32] and from the Epistemological Beliefs Assessment for Physical Sciences (EBAPS) [33], The validation studies of both surveys were carried out by their developers. Both surveys have been used in several studies to explore students' epistemologies [32,34–37].

MPEX-II is a discipline-specific instrument intended for high school and university physics courses. Building on previous work demonstrating dimensionality in students' epistemologies [2,18], MPEX-II was developed as a multidimensional survey based on Hammer's [4] dimensions for students taking a traditional physics course: pieces versus coherence, formulas versus concepts, and authority versus independence. We will call these dimensions coherence, concepts, and independence for short. Redish and Hammer [38] defined them as follows.

Coherence.—The degree to which the student sees physics knowledge as coherent and sensible as opposed to a bunch of disconnected pieces.

Concepts.—The extent to which the student sees concepts as the substance of physics as opposed to thinking of them as mere cues for which formulas to use. In other words, it is related to students' views about the content of physics knowledge as formulas or as concepts that underlie the formulas.

Independence.—The extent to which the student sees learning physics as a matter of constructing her own understanding rather than absorbing knowledge from authority.

MPEX-II contains two groups of items [39]. The first group consists of 25 Likert-type items on a five point scale. Items 26–32 are multiple choice, also scored on a five point scale, the score of which ranges from 1 point (a) to 5 points (e). Items 1 and 26 were not assigned to any dimension, though researchers can decide to include to gather additional information about students' views, such as views about group working. Table I shows dimensions, subdimensions, and the items falling into subdimensions and favorable responses in parenthesis.

To use MPEX- II in Turkey, it was converted into Turkish, in four steps. In the first step, the first author of this study translated it into Turkish. Second, this version

TABLE I.   Items falling into each dimension and subdimension of MPEX-II.

Dimensions	Items (favorable responses)
Coherence	3(D)[a], 4(A), 6(D), 8(D), 10(D), 13(D), 15(D), 19(D), 21(D), 23(D), 27(A,B), 28(A,C)
Concepts	5(D), 9(D), 16(D), 18(D), 19(D), 23(D), 24(D), 28(A,C), 30(D,E)
Independence	2(D), 7(A), 11(D), 12(D), 14(D), 17(A), 20(D), 22(D), 25(D), 29(D,E), 31(D,E), 32(A,B)

[a]Abbreviations: A, agree for 2-25 items; D, disagree for 2-25 items. For 27-32 items, A, B, C, D, and E indicate item alternative choices.

was examined by an instructor from the Basic English Department at her university, and the survey was revised accordingly. Third, this edited translation was reviewed by three instructors from the Faculty of Education who were familiar with the English version of MPEX-II. Their suggestions were synthesized and the changes were made. Finally, the Turkish version was examined by a high school teacher who teaches Turkish, to edit grammatical errors and check whether the items would be understandable to high school students.

III. PROCEDURE

The Turkish MPEX-II was administered to students by the first author. They took 25–30 min to complete it. The students indicated their gender on the survey.

Confirmatory factor analysis (CFA) was conducted using LISREL 8.30 [40] with SIMPLIS command language to test the hypothesized model for three-factor structure (corresponding to the *coherence*, *concepts*, and *independence* clusters) of the Turkish MPEX-II. Since the normality assumption was met, maximum likelihood estimation method was employed for CFA based on the covariance matrix. CFA is a structural equation modeling (SEM) technique used to investigate the degree to which the model-implied covariance-correlation matrix is equivalent to the empirical covariance-correlation matrix [41]. In SEM, model testing involves calculating the model-implied matrix and comparing it item by item with the observed matrix [42]. Because statistical analyses assume variables measured without error, multiple fit indices are helpful to evaluate to what extent the model fits the observed data [41]. In this study, χ^2/degrees of freedom (d.o.f.), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), standardized root-mean-square residual (SRMR), and root-mean-square error of approximation (RMSEA) were used for interpreting whether the model adequately fits the data [41,43]. χ^2 test statistic is used to test statistically whether the difference between the covariance matrix implied by the model and the empirical covariance matrix is equal to zero. That is, nonsignificant χ^2 indicates the model matches the data. However, when sample size increase, the possibility of getting significant χ^2 also increases, even for small differences [41–43]. Therefore, Jöreskog and Sörbom [43] suggest considering χ^2 as a measure of fit rather than test statistic and comparing its value to degrees of freedom. Thus, the ratio

χ^2/d.o.f. has been generally used to measure the fit [41,42]. Root-mean-square residual (RMR) is the square root of the mean of the squared discrepancies between the model-implied and empirical covariance matrix. Because this index depends on the scale of measurement of the variables, SRMR, which uses standardized instead of absolute discrepancies, has been introduced. RMSEA is similar to RMR based on residual analysis. GFI is calculated by taking a ratio of the sum of squared discrepancies to the observed variance [42]. AGFI is adjusted for degrees of freedom. GFI and AGFI indicate to what extent the model fits the data as compared to no model at all [43]. This study used the rule-of-thumb criteria for these goodness-of-fit indices as recommended by Schermelleh-Engel *et al.* [41]; see Table II.

Readers could wonder why we, unlike the original developers of MPEX and MPEX-II, investigated the psychometric properties of the survey—specifically, the reliability of the clusters, as described in the Results below—when the MPEX developers specifically avoided such analyses for principled reasons described below. We have three reasons. First and foremost, researchers working from the beliefs perspective may disagree with those principled reasons and therefore may want to know whether the Turkish MPEX-II is psychometrically reliable, so they can decide whether to use the survey and how strongly to interpret results. MPEX-II (or a translated version thereof) may be "reliable" even though its developers did not construct it to be. Second, from either a beliefs or a resources perspective, the clustering patterns of students' responses say something interesting about the survey items and/or students' epistemologies; see below for more details. Third, we wanted to use the Turkish MPEX-II to make the arguments presented in this paper, and to do so, we needed to conduct the "standard" analyses that survey creators perform.

TABLE II.   The rule of thumb criteria for goodness-of-fit indices. (Adapted from Schermelleh-Engel *et al.* [41].)

Fit measures	Good fit	Acceptable fit
χ^2/d.o.f.	$0 \leq \chi^2/\text{d.o.f.} \leq 2$	$2 < \chi^2/\text{d.o.f.} \leq 3$
RMSEA	$0 \leq \text{RMSEA} \leq 0.05$	$0.05 < \text{RMSEA} \leq 0.08$
SRMR	$0 \leq \text{SRMR} \leq 0.05$	$0.05 < \text{SRMR} \leq 0.10$
GFI	$0.95 \leq \text{GFI} \leq 1.00$	$0.90 \leq \text{GFI} < 0.95$
AGFI	$0.90 \leq \text{AGFI} \leq 1.00$	$0.85 \leq \text{AGFI} < 0.90$

IV. RESULTS

In this section, we first present the results of performing these statistical analyses described above on our Turkish MPEX-II data, and we then interpret those results through the lens of the beliefs perspective and modify the survey accordingly. Then, we reinterpret those same results from the resources perspective.

A. Applying the beliefs framework to the Turkish MPEX- II data

1. Reliability coefficients

Reliability coefficients were obtained using the Statistical Package for Social Studies (SPSS). Internal consistency reliability coefficients (Cronbach's alphas) for the Turkish MPEX-II and its three dimensions are presented in Table III.

It is expected that the value of Cronbach's alpha, indicating the internal consistency among the items, be at least 0.70 so that the survey can be considered reliable [44]. However, Cronbach's alphas of the three dimensions did not reach that threshold.

2. First confirmatory factor analysis

Table IV reports standardized factor loadings and t values of individual items. Most items loaded significantly ($p < 0.05$) on their hypothesized latent factors. However, five coherence items, items 10, 19, 23, 27, 28, one concepts item, item 28, and one independence item, item 32, had insignificant factor loadings.

So, although the fit indices were at an acceptable level ($\chi^2/\text{d.o.f.} = 4.03$, RMSEA = 0.078, SRMR = 0.073, GFI = 0.82, AGFI = 0.80), the hypothesized model for the three-factor structure of MPEX-II did not fit the data because of insignificant factor loadings.

Since beliefs perspective advocates consider students' epistemologies to be context-general, stable beliefs, they should be consistent across contexts, at least within a given discipline. Therefore, according to the beliefs perspective, different items all probing the same epistemological dimension should obtain similar results. It follows that, if an item purporting to probe the coherence dimension does not correlate with the other items probing that same dimension, something is wrong with the item. Hence, we should eliminate from the survey the items having insignificant correlation with other items in the same dimension.

TABLE III. Cronbach's alphas for the Turkish MPEX-II and its three dimensions.

Scales	α
MPEX-II	0.78
Coherence	0.56
Concepts	0.68
Independence	0.49

TABLE IV. Standardized factor loadings (t values) for items of the Turkish MPEX-II.

Items	Coherence	Concepts	Independence
Item 2			0.51 (9.50)
Item 3	0.49 (9.85)		
Item 4	0.21 (4.40)		
Item 5		0.68 (14.03)	
Item 6	0.51 (9.24)		
Item 7			0.22 (3.69)
Item 8	0.38 (7.05)		
Item 9		0.45 (8.72)	
Item 10	0.13 (1.94)[a]		
Item 11			0.45 (8.21)
Item 12			0.37 (6.16)
Item 13	0.55 (9.05)		
Item 14			0.33 (5.33)
Item 15	0.32(5.50)		
Item 16		0.74 (15.55)	
Item 17			0.41 (7.74)
Item 18		0.61 (10.10)	
Item 19	0.15 (0.88)[a]	0.37 (2.22)	
Item 20			0.49 (10.16)
Item 21	0.43 (7.49)		
Item 22			0.45 (7.98)
Item 23	0.22 (1.56)[a]	0.40 (2.92)	
Item 24		0.76 (16.34)	
Item 25			0.39 (7.32)
Item 27	0.05 (0.73)[a]		
Item 28	0.16 (0.87)[a]	0.01 (0.06)[a]	
Item 29			0.21 (3.86)
Item 30		0.13 (2.47)	
Item 31			0.16 (2.32)
Item 32			0.01 (0.16)[a]

[a]Factor loadings are insignificant ($p > 0.05$).

3. Second confirmatory factor analysis

After excluding the items with insignificant factor loadings and adding error variances between some items based on suggestions from modification indices, the model moderately fit to the data ($\chi^2/\text{d.o.f.} = 2.17$, RMSEA = 0.048, SRMR = 0.059, GFI = .91, AGFI = 0.89). All factor loadings are significant (see Table V).

Table VI reports Cronbach's alphas using items included in the second factor analysis. As can be seen from Table VI, all except for Cronbach's alpha of the coherence dimension increased after eliminating low correlated items. Cronbach's alpha of the coherence dimension decreased little. However, smaller Cronbach's alphas are common with short scales (fewer than 10 items) because it depends on the length of the scale [44,45]. Therefore, for short scales, mean interitem correlation (MIIC) is suggested for reliability concerns [44] since it is independent of the length of the scale [45]. The suggested value of MIIC is between 0.2 and 0.4 for reliability concerns [44]. Removing five items from the coherence dimension

TABLE V. Standardized factor loadings (*t* values) for items of the Turkish MPEX-II. (All factor loadings are significant ($p < 0.05$.)

Items	Coherence	Concepts	Independence
Item 2			0.50 (9.15)
Item 3	0.48 (9.46)		
Item 4	0.18 (3.51)		
Item 5		0.68 (14.18)	
Item 6	0.52 (9.19)		
Item 7			0.22 (3.79)
Item 8	0.39 (6.69)		
Item 9		0.45 (8.76)	
Item 11			0.46 (8.27)
Item 12			0.34 (5.71)
Item 13	0.52 (8.14)		
Item 14			0.23 (3.64)
Item 15	0.33 (5.45)		
Item 16		0.74 (15.58)	
Item 17			0.41 (7.72)
Item 18		0.60 (10.09)	
Item 19		0.48 (7.84)	
Item 20			0.51 (10.43)
Item 21	0.36 (6.13)		
Item 22			0.44 (7.67)
Item 23		0.59 (11.65)	
Item 24		0.75 (16.11)	
Item 25			0.39 (7.29)
Item 29			0.24 (4.47)
Item 30		0.12 (2.40)	
Item 31			0.15 (2.11)

significantly reduced the number of items. Thus, instead of comparing Cronbach's alpha value, we can compare MIIC values for the coherence dimension. It increased from 0.09 to 0.13. So, it can be said that reliability of the coherence dimension got higher as well, though that dimension is still not reliable by the usual threshold of alpha > 0.70 or MIIC > 0.2.

In summary, eliminating the "bad" items from the Turkish MPEX resulted in an instrument that is still only marginally reliable; factor analysis and calculation of Cronbach's alphas indicate that the abridged survey is reliable overall and in the concepts dimension, but not in the coherence or independence dimensions. From the beliefs perspective, the survey is therefore of only limited use.

TABLE VI. Cronbach's alphas for the Turkish MPEX-II and its three dimensions after the second factor analysis.

Scales	α
MPEX-II	0.80
Coherence	0.51
Concepts	0.71
Independence	0.52

4. Next steps, according to the beliefs perspective

For a researcher working within the beliefs perspective, the next steps would be to find the items within the coherence and independence clusters that correlate least strongly with the rest, and then refine, replace, or delete those items. Then, the psychometric tests discussed above would be run on the revised survey. This iterative refinement process would continue until those clusters reached minimum psychometric thresholds of reliability.

B. Applying resource framework to the Turkish MPEX- II data

The developers of MPEX-II did not conduct a factor analysis to test whether the items within a given subscale all probe the same beliefs, for a principled reason [33]. According to the resources perspective, students' epistemological knowledge consists of context sensitive fine-grained resources. Therefore, low correlations among some items within a given dimension do not necessarily indicate that some of the items are "bad"; the low correlations could indicate that different contextual cues in different items triggered different sets of epistemological resources [19]. For this reason, the dimensions of an epistemology survey such as MPEX-II should be viewed as targets of instruction rather than as beliefs corresponding to stable cognitive structures [26,46].

1. Potential invalidity of psychometric "unreliability"

To flesh out this argument, we now take a closer look at some of the items that got removed from MPEX-II after the first factor analysis. Item 27, presented below, is one such example. Students' answers to this item correlated only weakly with their answers to other items in the coherence dimension. According to the beliefs perspective, those low correlations indicate that the question is problematic. Specifically, most students' answers to item 27 indicated a favorable view about the coherence of physics knowledge, but many of those students' displayed neutral or unfavorable ("piecemeal knowledge") views in their responses to other coherence items, resulting in low correlations between item 27 and some of those other items. According to the beliefs perspective, each student has a certain belief about the extent to which physics knowledge is coherence; therefore, the fact that many students' responses to item 27 "disagree" with their responses to other coherence items indicates that item 27 is unreliable at revealing students' actual beliefs about coherence. Hence, item 27 gets removed from the survey.

In contrast, according to the resources perspective, many students do not *have* stable beliefs about the coherence of physics knowledge; some contexts might trigger more coherence-seeking epistemological stances while other contexts might trigger more "piecemeal knowledge" stances. Therefore, a survey item's insignificant factor

loading, or a cluster's low alpha, could indicate actual context-dependent variability in students' epistemologies rather than unreliability of survey items. Of course, some of the "unreliable" might actually be invalid, in the sense of failing to probe what it was intended to probe. Our point is that psychometric analyses alone cannot tell us which "unreliable" items are really invalid versus which items uncover real variability in student epistemologies.

To illustrate this point, we will contrast item 27, whose factor loading was statistically insignificant, to item 6, whose factor loading was significant. Both of those coherence items probe students' views about whether physics knowledge forms a coherent, interconnected whole or consists of disconnected pieces.

Item 27: In the following question, you will read a short discussion between two students who disagree about some issue. Then you'll indicate whether you agree with one student or the other.

Tracy [47]: A good physics textbook should show how the material in one chapter relates to the material in other chapters. It shouldn't treat each topic as a separate "unit," because they're not really separate.

Carissa: But most of the time, each chapter is about a different topic, and those different topics don't always have much to do with each other. The textbook should keep everything separate, instead of blending it all together.

With whom do you agree? Read all the choices before choosing one.

(a) I agree almost entirely with Tracy.

(b) Although I agree more with Tracy, I think Carissa makes some good points.

(c) I agree (or disagree) equally with Carissa and Tracy.

(d) Although I agree more with Carissa, I think Tracy makes some good points.

(e) I agree almost entirely with Carissa.

Item 6: Knowledge in physics consists of many pieces of information, each of which applies primarily to a specific situation.

61% of the students gave favorable responses to item 27 [choices (a) or (b)], while only 27% percent of the students gave favorable responses to item 6 (strongly or somewhat disagree). Why did so many students disagree with themselves across these two questions?

Our speculative, partial answer stems in part from the first author's research into how Turkish high school physics students approach their learning [48,49]. Textbooks are rarely used, which means item 27 is unlikely to get students thinking about their classroom experiences. Furthermore, item 27 asks students to consider a debate instead of giving a quick, gut response. For these two reasons, item 27 might tap into the more "philosophical" side (as opposed to the classroom survival side) of students' epistemologies. In this more reflective, philosophical mode, students may be more likely to think of physics knowledge as ultimately coherent; as Hammer [4] found, even students who learn physics as a bunch of pieces of information, and do not think they can do otherwise, often know that experts see the knowledge as coherent. These factors could help explain why item 27 elicited more sophisticated responses than did other coherence items.

In contrast, the brevity of item 6—and its placement in a block of other such items on the survey—invites a quicker gut response, and the mention of "applies primarily to a specific situation" could tap into the ways that Turkish high school students get trained for the high-stakes college entrance exam: by solving hundreds of problems, each of a particular identifiable kind (circular motion, blocks on ramps, etc.), trying to improve their speed but not their deeper understanding. If item 6 indeed triggers students to think about this approach to studying, it is no wonder that only 27% give the favorable response (denying that physics consists of situation-specific "pieces" of information). Indeed, in another study [48] of a subset of the Turkish students who took MPEX-II, we asked students to consider a hypothetical student, Arzu, who wants to understand physics well but who does not need to take the standardized college entrance exams that most Turkish physics students take. Writing about how Arzu should study, many students said that she should seek connections between different topics, a stance that would lead to a favorable response to MPEX-II item 27. In contrast, writing about their own study habits, many students emphasized the importance of learning to quickly solve many different types of problems—a stance that could lead to an unfavorable response to item 6. So, students' "inconsistent" responses to items 27 and 6 could result from item 27 tapping into the more reflective, idealistic "Arzu" side of their epistemologies and item 6 tapping into the more practical, classroom-cued side of their epistemologies.

We have just given a plausibility argument that the disparity in many students' responses to item 6 versus item 27 reflects context-specific variability in their epistemologies: they possess epistemological resources for viewing physics as disconnected pieces of information, cued perhaps by exam preparation strategies, and they also possess resources for viewing physics knowledge as coherent, cued perhaps by contexts of philosophical reflection. According to the resources perspective, neither of these epistemological stances reflects a student's one "real" epistemology. Instead, both stances reflect local coherences in the students' epistemic cognition, and hence, the psychometric unreliability of one of those two items (item 27) is not a reason to drop that item from the survey. Indeed, removing item 27 would deprive researchers of information about what kinds of contextual cues tend to elicit the more sophisticated pockets of students' epistemologies of physics.

2. What psychometric analyses can tell us, from a resources perspective

In previous work, the second author has resisted psychometric analyses of epistemology survey items for the reasons discussed above. However, such analyses *do* serve a purpose, even from the resources perspective, by flagging items that might actually be bad in the sense of not probing the intended dimension of epistemology. Again, psychometric analyses alone cannot tell us whether the item is actually problematic or whether it merely reveals variability in students' epistemologies. Further study of the flagged items is needed to decide, and in some cases, the researchers may conclude the item actually is a bad fit for the survey.

For instance, consider item 32.

Item 32: Several students are talking about group work.
Carmela: "I feel like explaining something to other people in my group really helps me understand it better."
Juanita: "I don't think explaining helps you understand better. It's just that when you can explain something to someone else, then you know you already understood it."
With whom do you agree? Read all the choices before choosing one.
(a) I agree almost entirely with Carmela.
(b) Although I agree more with Carmela, I think Juanita makes some good points.
(c) I agree (or disagree) equally with Juanita and Carmela.
(d) Although I agree more with Juanita, I think Carmela makes some good points.
(e) I agree almost entirely with Juanita.

This item passed our interview-based validity testing; students' answers do indeed reflect their views about the value of explaining things to other people, as intended. But factor analysis revealed that this item correlates poorly with other items in the independence dimension; a disproportionately high percentage of students gave favorable responses. This psychometric problem caused us to rethink the item, and we now suspect that it might not probe students' (context-dependent) views about constructivist versus transmissionist learning. A student can agree with Carmela not because she thinks learning involves constructing one's own understanding, but rather, because she thinks that *rehearsing* her understanding helps reinforce it, even if she thinks she absorbed rather than constructed her original understanding.

3. Next steps, from the resources perspective

Interviews targeting item 32, where the researcher asks students to explain their answer and asks a lot of follow-up questions about the reasons a student might or might not benefit from explaining her answer, can help us pin down whether this "rehearsal" explanation explains students'

agreement with Carmela, or whether this group-work-situated item actually taps into students' constructivist ideas. Our point here is that the psychometric analyses flagged this item for reexamination, and productively so.

Regarding items 6 and 27, the next step would be interviews in which we ask students to explain the reasoning behind their answers. We would code their responses for whether they are predominantly about (i) the student's "general" views of what physics is and what it means to learn physics versus (ii) their own study habits and needs. Confirmation of our hypothesis that the ratio of type (i) to type (ii) response is higher for item 27 than for item 6 would be evidence that those two items tap into different sides of students' epistemologies, as discussed above.

In brief, from the resources perspective, psychometrics alone cannot deem an item "unreliable." But it can help researchers rethink and retest items. In some cases, a plausible—and ultimately testable—context-based account can explain why students responded differently to that item than they did to other items in the same dimensions. In other cases, the researchers need to rethink what that item is actually probing.

V. CONCLUSIONS AND DISCUSSION

In this study, we used standard psychometric techniques to characterize and improve the reliability of the Turkish version of the MPEX-II survey of students' epistemologies and expectations, administered to several hundred Turkish high school physics students. Even after eliminating "bad" items, two of the three dimensions probed by MPEX-II did not correspond to psychometrically reliable factors; the Cronbach's alphas were below the usual threshold of 0.70 because some of the items within the dimension correlated too weakly with other items in the dimension. When interpreted in terms of the beliefs perspective, this unreliability is taken to mean that the survey fails to reveal students *actual* beliefs within the two dimensions, and the next step is to refine or replace the weakly correlated survey items. Underlying this conclusion is the beliefs-based assumption, often implicit in purely psychometric analyses and discussions, that students indeed have stable epistemological beliefs that a sufficiently good survey can read out. Therefore, if a student "disagrees with herself" when answering different items within a given dimension, it is because one or more of the items is a bad probe of the student's belief.

The resources perspective supplies a different interpretation of psychometric unreliability. By this perspective, students' epistemological views are context dependent; different contextual cues embedded in different survey items can trigger different sets of epistemological resources. Therefore, when students' answers to two different items within a given dimension do not correlate, it can be because one of the items does not probe the desired dimension—but it can also be because of actual variability

in students' epistemologies. In that case, eliminating one of the items to increase reliability could actually decrease the validity and usefulness of the survey, by hiding from the view the rich context dependencies in students' response patterns and by cutting off researchers from clues about which kinds of contextual cues tend to evoke the more sophisticated versus the less sophisticated pockets of students' epistemological resources. By this account, researchers should consider retaining items and survey clusters that are psychometrically unreliable, with the caveat, of course, that the survey dimensions are interpreted as targets of instruction, not as stable beliefs.

From the resources perspective, the next step for a researcher who finds such inconsistencies in students' responses is to formulate plausible context-based hypotheses about those inconsistencies and to test those hypotheses using interviews or other methods that probe students' reasoning more deeply than a forced-response survey can do.

In conclusion, this study explored how different epistemological frameworks interpret psychometric analyses of the Turkish MPEX-II data. Our broader point is that psychometrically driven heuristics for survey design and interpretation are not ''nonpartisan'' best practices; they are heavily theory laden. Specifically, they reflect the beliefs perspective. We showed that the beliefs and resources frameworks provided different interpretations of the psychometric analyses, leading to different conclusions about how the survey results should be interpreted and what next steps should be taken to improve the survey and/or better understand its results.

ACKNOWLEDGMENTS

[1] In the literature, there is a discussion about whether the definition of personal epistemology should include students' views about learning. See A. Elby, Defining personal epistemology: A response to Hofer & Pintrich (1997) and Sandoval (2005), J. Learn. Sci. **18**, 138 (2009). and Ref [2]. In this study, students' understandings of the nature of learning are considered as part of their personal epistemologies

[2] B. K. Hofer and P. R. Pintrich, The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning, Rev. Educ. Res. **67**, 88 (1997).

[3] B. K. Hofer, in *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, edited by B. K. Hofer and P. R. Pintrich (Lawrence Erlbaum, Mahwah, NJ, 2002), pp. 3–14.

[4] D. Hammer, Epistemological beliefs in introductory physics, Cogn. Instr. **12**, 151 (1994).

[5] L. Lising and A. Elby, The impact of epistemology on learning: A case study from introductory physics, Am. J. Phys. **73**, 372 (2005).

[6] D. B. May and E. Etkina, College physics students' epistemological self-reflection and its relationship to conceptual learning, Am. J. Phys. **70**, 1249 (2002).

[7] M. P. Ryan, Monitoring text comprehension: Individual differences in epistemological standards, J. Educ. Psychol. **76**, 248 (1984).

[8] M. Schommer, A. Cruse, and N. Rhodes, Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so, J. Educ. Psychol. **84**, 435 (1992).

[9] N. B. Songer and M. C. Linn, How do students' views of science influence knowledge integration?, J. Res. Sci. Teach. **28**, 761 (1991).

[10] K. Hogan, Relating students' personal frameworks for science learning to their cognition in collaborative contexts, Sci. Educ. **83**, 1 (1999).

[11] S. A. Rosenberg, D. Hammer, and J. Phelan, Multiple epistemological coherences in an eighth-grade discussion of the rock cycle, J. Learn. Sci. **15**, 261 (2006).

[12] W. A. Sandoval, Understanding students' practical epistemologies and their influence on learning through inquiry, Sci. Educ. **89**, 634 (2005).

[13] C.-C. Tsai, An analysis of scientific epistemological beliefs and learning orientations of Taiwanese eighth graders, Sci. Educ. **82**, 473 (1998).

[14] L. Louca, A. Elby, D. Hammer, and T. Kagey, Epistemological resources: Applying a new epistemological framework to science instruction, Educ. Psychol. **39**, 57 (2004).

[15] P. M. King and K. S. Kitchener, *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults* (Jossey, San Francisco, 1994).

[16] D. Kuhn, R. Cheney, and M. Weinstock, The development of epistemological understanding, Cognit. Dev. **15**, 309 (2000).

[17] C. L. Smith and L. Wenk, Relations among three aspects of first-year college students' epistemologies of science, J. Res. Sci. Teach. **43**, 747 (2006).

[18] M. Schommer, Effects of beliefs about the nature of knowledge on comprehension, J. Educ. Psychol. **82**, 498 (1990).

[19] D. Hammer and A. Elby, in *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, edited by B. K. Hofer and P. R. Pintrich (Lawrence Erlbaum, Mahwah, NJ, 2002), pp. 169–190.

[20] L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics education research, Am. J. Phys. **67**, 755 (1999).

[21] B. K. Hofer, Dimensionality and disciplinary differences in personal epistemology, Contemp. Educ. Psychol. **25**, 378 (2000).

[22] M. Schommer and K. Walker, Are epistemological beliefs similar across domains?, J. Educ. Psychol. **87**, 424 (1995).

[23] M. M. Buehl, P. A. Alexander, and P. K. Murphy, Beliefs about schooled knowledge: Domain specific or domain general?, Contemp. Educ. Psychol. **27**, 415 (2002).

[24] P. Op'T Eynde, E. De Corte, and L. Verschaffel, Epistemic dimensions of students' mathematics-related belief systems, Int. J. Educ. Res. **45**, 57 (2006).

[25] J. Leach, R. Millar, J. Ryder, and M. G. Sere, Epistemological understanding in science learning: The consistency of representations across contexts, Learn. Instr. **10**, 497 (2000).

[26] W. M. Roth and A. Roychoudhury, Physics students' epistemologies and views about knowing and learning, J. Res. Sci. Teach. **31**, 5 (1994).

[27] W. A. Sandoval and K. Morrison, High school students' ideas about theories and theory change after a biological inquiry unit, J. Res. Sci. Teach. **40**, 369 (2003).

[28] A. A. diSessa, Towards an epistemology of physics, Cogn. Instr. **10**, 105 (1993).

[29] A. Elby and D. Hammer, in *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice*, edited by L. D. Bendixen and F. C. Feucht (Cambridge University Press, Cambridge, England, 2010), pp. 409–434.

[30] D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. Mestre (Information Age Publishing, Greenwich, CT, 2005), pp. 89–120.

[31] A. Elby, T. McCaskey, R. Lippmann, and E. F. Redish, http://www.physics.umd.edu/perg/tools/attsur.htm.

[32] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66**, 212 (1998).

[33] B. White, A. Elby, J. Frederiksen, and C. Schwarz, The epistemological beliefs assessment for physical science, in Proceedings of the American Education Research Association Conference, Montreal, 1999 (unpublished).

[34] C. H. Crouch and E. Mazur, Peer instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

[35] A. Elby, Helping physics students learn how to learn, Am. J. Phys. **69**, S54 (2001).

[36] R. Pulmones, Linking students' epistemological beliefs with their metacognition in a chemistry classroom, Asia-Pac. Educ. Researcher **19**, 143 (2010).

[37] M. Sahin, Exploring university students' expectations and beliefs about physics and physics learning in a problem-based learning context, Eurasia J. Math. Sci. Technol. Educ. **5**, 321 (2009).

[38] E. F. Redish and D. Hammer, Reinventing college physics for biologists: Explicating an epistemological curriculum, Am. J. Phys. **77**, 629 (2009).

[39] All survey items are available on the public web page of University of Maryland Physics Education Research Group. The survey can be retrieved from http://www.physics.umd.edu/perg/tools/MPEX-II.pdf.

[40] K. G. Jöreskog and D. Sörbom, computer code LISREL 8.30, Scientific Software International, Inc., Chicago, 1999.

[41] K. Schermelleh-Engel, H. Moosbrugger, and H. Muller, Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures, Meth. Psychol. Res. Online **8**, 23 (2003).

[42] E. K. Kelloway, *Using LISREL for Structural Equation Modeling* (Sage Publications, London, 1998).

[43] K. G. Jöreskog and D. Sörbom, LISREL 8: *Structural Equation Modeling with SIMPLIS Command Language* (Scientific Software International, Chicago, 1993).

[44] J. Pallant, SPSS Survival Manual: *A Step by Step Guide to Data Analysis Using SPSS for Windows (Versions 10 and 11)* (Open University Press, Maidenhead, England, 2001).

[45] S. R. Briggs and J. M. Cheek, The role of factor analysis in the development and evaluation of personality scales, J. Pers. **54**, 106 (1986).

[46] A. Elby, http://www2.physics.umd.edu/~elby/EBAPS/home.htm.

[47] Names of students were changed with Turkish girl names in the Turkish MPEX.

[48] S. Yerdelen-Damar and A. Elby (to be published)

[49] S. Yerdelen-Damar and A. Elby, The effect of high-stake testing on students' learning approaches in physics, in Proceedings of the Physics Education Research Conference, Omaha, NE, 2011 (unpublished).