En Búsqueda de Evidencia Convergente de Creencias Epistemológicas: una Encuesta Novel

Jeremy E. Briell¹, Jan Elen², and Geraldine Clarebout³

¹Jeremy E. Briell, Center for Instructional Psychology and Technology.
²Jan Elen, Center for Instructional Psychology and Technology.
³Geraldine Clarebout, Centre for Medical Education. University of Leuven

Belgium

Correspondence: Jeremy E. Briell, Center for Instructional Psychology and Technology, University of Leuven, Van Den Heuvel-Instituut, Dekenstraat 2, pb 3773 B - 3000 Leuven. Belgium. E-mail: jeremy.briell@student.kuleuven.be

© Education & Psychology I+D+i and Editorial EOS (Spain)
Resumen

Introducción. En este trabajo se explica el desarrollo y prueba de un nuevo instrumento de papel y lápiz diseñado para evaluar y apoyar las inferencias de cinco dimensiones de creencias epistemológicas sobre la base de tres tipos de pruebas. Se prevé que los tres tipos de evidencia que convergen en un solo punto de vista teórico, permitiendo inferencias mejores soportadas que las medidas de las encuestas existentes.

Método. El cuestionario que evidencia de múltiples fuentes de las creencias epistemológicas (MESSEB) se puso a prueba con 12 estudiantes de nivel universitario de Hong Kong (19 a 34 años de edad, M = 24.9, SD = 4.2).

Resultados. Los resultados indican que las puntuaciones de los tres tipos de pruebas no convergen. En su lugar, un resumen de los resultados tuvo que ser determinado en base a las puntuaciones medias para cada dimensión y para todas las dimensiones combinadas. Como grupo, los participantes puntuaron moderadamente en todas las dimensiones combinadas, puntuando peor en la organización del conocimiento, la justificación de las dimensiones del conocimiento y mejor en la estabilidad de la dimensión del conocimiento.

Discusión y Conclusiones. Se examinan tres explicaciones de la variabilidad de este tipo pruebas: (a) la MESSEB produce adecuadamente las puntuaciones previstas, (b) los tipos de pruebas no son apropiadas para inferir las creencias epistemológicas previstos, y (c) las creencias epistemológicas se miden según lo previsto, pero variación es genuino y típica. Se analizan las implicaciones de cada explicación, junto con las posibilidades de abordarse en futuras investigaciones.

Palabras clave: epistemología personal, creencias epistemológicas, creencias, diseño de cuestionario.
Seeking Convergent Evidence of Epistemological Beliefs: A Novel Survey

Abstract

Introduction. This paper explains the development and testing of a novel paper-and-pencil measure designed to support inferences of five epistemological belief dimensions based on three forms of evidence. The three evidence types are anticipated to converge at a single theoretical level, permitting better-supported inferences than existing survey measures.

Method. Data were collected from 12 university-level students from Hong Kong (19 to 34 years of age, M = 24.9, SD = 4.2). Participants completed a survey designed to capture three different sources of evidence of epistemological beliefs: epistemological judgments, intentions, and explicit beliefs. A theoretical level (i.e., dualism, relativism discovered, and contextual relativism) is assigned for each evidence type for each of five dimensions. The theoretical-level assignments across the three evidence sources are then analyzed for convergence.

Results. Results indicate the three assigned theoretical levels for each dimension did not converge. Therefore, a comprehensive theoretical level per dimension could not be given. Instead, summative scores could to be determined based on average performance per dimension and for all dimensions combined. As a group, participants scored moderately on all dimensions combined, performing worst on the Organization of knowledge and Justification of knowledge dimensions and best on the Stability of knowledge dimension.

Discussion and Conclusions. Three explanations for evidence type variability are discussed: a) the MESSEB is incorrectly eliciting the intended evidence, (b) the types of evidence are not appropriate for inferring the intended epistemological beliefs, and (c) epistemological beliefs are being measured as intended, but variation is genuine and typical. The ramifications of each explanation are discussed along with possibilities to address them in future research.

Keywords: Personal Epistemology; Epistemological Beliefs; Beliefs; Survey Design

Received: 04/28/13 Initial acceptance: 05/15/13 Final acceptance: 07/20/13
Introduction

Epistemological beliefs are highly relevant to educators because they have long been thought to have powerful influences in mediating students’ thinking and learning (e.g., King & Kitchener, 1994; Kuhn, 1991; Schommer, 1990). Nevertheless, they represent a construct that is known to be difficult to measure, with several popular instruments receiving sharp criticism (see Clarebout, Elen, Luyten, & Bamps, 2001; see also Maggioni, Riconscente, & Alexander, 2006). This paper details the development and testing of a unique paper-and-pencil measure of epistemological beliefs. We begin by explaining our theoretical underpinnings, including how we define epistemological beliefs and how we believe they should be measured, which serves as a guide for the design of instrument parts and items (see Hinkin; 1998; Schwab, 1980).

Conceptualization of Epistemological Beliefs

Epistemological beliefs are the abstract beliefs of lay folk that address questions relevant to professional epistemologists, typically about the nature of knowledge and knowing. (see Briell, Elen, Verschaffel, & Clarebout, 2011; see also, Hofer & Pintrich, 1997). Although, epistemological beliefs can be conceived as pertaining to particular cognitive ranges - domain-general beliefs (i.e., beliefs about the general nature of knowledge and knowing), domain-specific beliefs (e.g., beliefs about the nature of mathematics), topic-specific beliefs (e.g., beliefs about the teaching of conceptual learning), or even context specific, known as epistemological resources (see Hammer & Elby, 2002) - our interest lies in beliefs about the nature of knowledge and knowing on a general level, which is in line with longstanding perspectives on the construct (see, for example, King & Kitchener, 1994 or Perry, 1970).

Early stage models, such as Epistemological Reflection (Baxter Magolda, 1992), Women’s Ways of Knowing (Belenky, Clinchy, Goldberger, & Tarule, 1986), the Reflective Judgment Model (King & Kitchener, 1994, 2004), and the Perry scheme (Perry, 1970, 1981) show that epistemological beliefs exist at different qualitative levels, typically regarded as a developmental pathway. These early models have tended to portray a “parallel trajectory, but significant points of distinction as well” (Hofer, 2001, p. 356). The Perry scheme serves as a good reference. Like most stage models, it portrays three basic levels of development: dualism (Positions 1-2) represents a dichototic view of knowledge as good or bad, right or
wrong, true or false. Authorities know the right answers and it is their job to impart that knowledge to students. The first major shift in the model is to a beginning awareness of relativism, referred to as relativism discovered (Positions 3-4). At this level, the awareness of uncertainty about what is good, right, and true can solidify into a certainty that these things can never be known for sure, that there are only arbitrary grounds for determining rightness. The final shift in understanding is to contextual relativism (Positions 5-9), awareness that:

- Diversity of opinion, values, and judgment [are] derived from coherent sources, logics, systems, and patterns allowing for analysis and comparison. Some opinions may be found worthless, while there will remain matters about which reasonable people reasonably disagree. Knowledge is qualitative, dependent on contexts. (Perry, 1981, p. 80 [italics in original text])

The Perry scheme and other similar developmental models portray epistemological beliefs as a unitary construct, that is, a set of beliefs that coherently exist at a single theoretical level. Schommer (1990), followed by Hofer and Pintrich (1997), discerned prominent developmental models could be teased apart into certain conceptual facets or dimensions. These efforts provide essential detail about specifically what epistemological-belief dimensions are relevant. Schommer (1990) theorized dimensions operate as a system of relatively independent beliefs, and, hence, could develop asynchronously. Despite an abundance of research into particular dimensional beliefs, asynchronous development has not been formally tested (Schommer-Aikins, 2004). Dimensional models have the clear advantage of offering detail about what epistemological-belief dimensions are pertinent and not insisting that they must congeal into a single coherent stage.

In their extensive review of the literature, Hofer and Pintrich (1997) proposed four dimensions: Certainty of knowledge (ranging from conceptions of knowledge being fixed to being tentative and evolving), Simplicity of knowledge (ranging from conceptions of knowledge as discrete pieces of information to highly interrelated concepts), Source of knowledge (ranging from conceptions of knowledge being derived from external authorities to conceptions of self as knower), and Justification for knowing (“how knowledge claims are evaluated, including the use of evidence, the use they make of authority and expertise, and their evaluation of experts” [p. 120]). This conceptualization contrasts Schommer’s (1990) earlier portrayal of the construct, which included dimensions regarding the speed of learning and innateness of learning ability.
We largely agree with the dimensions by Hofer and Pintrich (1997), because they are common to psychological and philosophical treatments of epistemology. Restricting epistemological-belief dimensions to those that address questions relevant to professional epistemologists is imperative to construct clarity (see Briell et al., 2011). However, we do have some differences in our articulation of the dimensions proposed by Hofer and Pintrich. Our first difference is one of nomenclature. The Simplicity of knowledge is more precisely referred to as the Organization of knowledge, because it refers to how knowledge is organized in the mind. Second, the Justification of knowledge as defined by Hofer and Pintrich concerns how particular knowledge claims are justified instead of abstract beliefs about the nature of justified knowledge. Although abstract beliefs about what constitutes justified knowledge are often inferred from how particular knowledge claims are justified, we believe it is important to make a conceptual distinction (see Briell et al., 2011). Hence, we define epistemological beliefs about the justification of knowledge as abstract beliefs about what - on a general level in this manuscript - constitutes justified knowledge. The Reflective Judgment Model and the Perry scheme indicate this dimension could range from a belief that knowledge requires no justification to a belief that proper evidence and reasons justify knowledge to others and enable deeper understandings. Finally, Certainty of knowledge, as written by the reviewers, appears to merge two distinct beliefs: (a) the certainty of knowledge in relation to beliefs about the stability of knowledge over time and (b) the certainty of knowledge in relation to beliefs about the diversity of problem resolutions. Hence, this dimension is better represented as two dimensions: Stability of knowledge (ranging from a belief that knowledge exists as a fixed entity in time to it being something that evolves continually) and Diversity of knowledge. Based on the Perry scheme, the latter would range from a belief that every problem has a single, correct answer that can be known with certainty to a belief in persistent uncertainty since many problems have multiple right answers that are dependent on contexts, subjective qualities of the knower, logic systems, and methods of analysis.

The Challenge of Measuring Epistemological Beliefs

Though epistemological beliefs research has been ongoing since the 1950’s and 60’s with Perry’s original work (Perry, 1970) and has become increasingly prolific, personal epistemological researchers are continually vexed with the fundamental issue of aptly and accurately measuring epistemological beliefs (see Briell et al., 2011).
A construct is a representation of something that exists in the subject’s mind; as its level of abstractness increases, it becomes more resistive to being measured (Nunnally, 1976). The highly abstract nature of epistemological beliefs makes it understandable why they are especially difficult to measure. Measurement challenges are confounded by the fact that epistemological beliefs – like other abstract beliefs - are normally regarded as implicitly held in the mind. Thus, inferential methods are requisite and must depend on accurately interpreting the visible manifestations of the underlying cognitive structures. Understandably, this is an obscure task and the many available instruments testify to the reality that researchers have yet to agree on the best way to infer epistemological beliefs. The review by Briell et al. (2011) indicates four basic methods for measuring epistemological beliefs: (a) inference from related beliefs, (b) inference from explicit epistemological beliefs, (c) inference from epistemological judgments, and (d) inference from one’s meaning making of issues of perceived importance. Schommer’s (1990) epistemological belief questionnaire is a prominent example of inference from related beliefs. Participants rate a level of agreement to beliefs thought to be indicative of the intended epistemological belief dimensions. In a second category, some authors ask more direct questions to participants about their epistemological beliefs (Berthelsen, Brownlee, & Boulton-Lewis, 2002; Burns & Bond, 2004; Maggioni et al., 2006). These studies still rely on inference, if it is accepted that there is more to one’s epistemological beliefs than might be revealed in explicit accounts (see Buehl & Alexander, 2006). Representing the third category, The Reflective Judgment Interview (King & Kitchener, 1994) allows inferences to be made based on participants’ judgments and justifications when confronted with competing expert assertions (referred to as epistemological judgments by Briell et al.). Indicative of the fourth category, Baxter Magolda (2004) makes inferences based on her participants’ meaning making of important life events. Briell et al. argue that each approach is fallible and that it is impossible to know “whether various measures are equally effective at accessing the same construct or even if they are accessing the same construct” (p. 28). They conclude that the way forward lies in identifying what constitutes a suitable inference, to which they further argue that inferences should be bound by multiple sources of evidence.

One articulation of what constitutes a suitable inference of belief has been championed by Pajares (1992), who summarized the measurement of belief as something “that can only be inferred from what people say, intend, and do” (p. 316). In his view, an accurate inference depends upon congruence between “belief statements, the intentionality to behave in a
predisposed manner, and the behavior related to the belief in question (p. 326). According to Pajares the three evidence types should correspond to a certain belief. For personal epistemological researchers, this means inferences might be bettered if drawn from participants’ stated beliefs about the nature of knowledge and knowing in conjunction with their intentions and actions toward the nature knowledge and knowing.

Aims

The main purpose of the pilot study reported here is to determine whether multiple types of data (as recorded by the designed survey described below) converge at a single theoretical level. If convergent evidence is achieved, theoretical levels could be assigned with a level of confidence.

1) Research questions:

1. Do results provide convergent evidence to support theoretical level assignments for epistemological belief dimensions (i.e., do the three forms of data that are recorded concur at a single theoretical level for each dimension)?
2. How do participants as a group perform on specific dimensions?
3. How do participants as a group perform on dimensions combined?

2) Hypotheses. We expect results from the different evidence sources will provide unanimous support for a single-theoretical-level assignment per dimension per participant. If this occurs, we will feel justified in making a comprehensive inference for the respective dimension.

Our university-aged sample ranges in education level and discipline, so as a group we would expect moderate performance on specific dimensions and overall performance.

Method

Participants

Participants were sought from Hong Kong (where the first author resides) via a snowball approach. Acquaintances of the first author helped select potential volunteers. Potential volunteers having met the criteria of being English proficient and a university student were asked to participate and to recommend other possible volunteers. Potential volunteers were contacted by email. The purpose of pilot was explained and their
participation was requested. Those opting to volunteer were reminded that participation would be kept confidential.

Participants were 12 volunteers ranging in age from 19 to 34 years ($M = 24.9, SD = 4.2$). Eight participants were male and four female. Education level varied from first year undergraduate to doctoral level, all from a range of disciplines. Participants have self-stated English proficiency.

**Instrument and Procedure**

1) *Survey design and pretesting.* The above theoretical foundation guides our development of The Multiple Evidence Source Survey of Epistemological Beliefs (MESSEB) (see Hinkin; 1998; Schwab, 1980). Namely, we set out with the goal of creating an instrument that would permit inferences of the epistemological beliefs of university-aged participants based on multiple sources of information in a way considerate of the ideal endorsed by Pajares (1992). However, we wanted an instrument that could be efficiently and systematically scored for good-sized population samples. Our first and second goals were only compatible as much as we limited the first goal to what would be possible with a paper-and-pencil measure. Our third goal was that the instrument would assess epistemological beliefs dimensions - that have a history of relevance in epistemological beliefs research - according to three qualitatively different levels. Our fourth goal is that dimensions are to be understood on a domain-general level, that is, the instrument is to address questions about the nature of knowledge and knowing in general.

Hofer and Pintrich (1997) have done a tremendous service in suggesting domain-general dimensions that are pertinent to philosophical and psychological treatments of epistemology, which we reinterpret above as the following five dimensions: *Organization of knowledge, Justification of knowledge, Source of knowledge, Stability of knowledge,* and *Diversity of knowledge.* Developmental models suggest three basic qualitatively different levels to epistemological beliefs (e.g., dualism, relativism discovered, contextual relativism). There is no available scheme that individually models all of these belief dimensions according to qualitatively different levels. This meant a detailed theoretical model had to be created. Hence, we constructed a model by referencing the Perry scheme, the Reflective Judgment Model, and Hofer and Pintrich’s (1997) review. Beliefs for each dimension indicative of
dualism, relativism discovered, and contextual relativism were theoretically decided. Table 1 below is the result of this effort.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Dualism</th>
<th>Theoretical Level of Sophistication</th>
<th>Contextual Relativism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source of knowledge</td>
<td>Knowledge is transmitted; the role of self is limited to receiving</td>
<td>Only the significance of self in the acquisition of knowledge is evident</td>
<td>The role of self and intentional instruction are both important in the acquisition of knowledge</td>
</tr>
<tr>
<td>Justification of knowledge</td>
<td>One can rightly say, “I know” only when there is proven correspondence to an objective reality</td>
<td>Facts and reasons do not enhance understandings or justify one’s claims to others</td>
<td>Evidence and reasons enables deeper/critical understandings and justifies one’s claims.</td>
</tr>
<tr>
<td>Organization of knowledge</td>
<td>Knowledge in the mind exists as well organized, but isolated units of information</td>
<td>Knowledge in the mind exist as disorganized units of information</td>
<td>Knowledge in the mind exists as interconnected units of information</td>
</tr>
<tr>
<td>Diversity of knowledge</td>
<td>Every problem has a single, correct answer</td>
<td>Correct/incorrect is <em>largely subjective</em></td>
<td>Correct answers are contingent upon <em>subjective and contextual/objective factors</em></td>
</tr>
<tr>
<td>Stability of knowledge</td>
<td>Knowledge is generally static over time</td>
<td>Knowledge changes drastically over time</td>
<td>Knowledge changes moderately with time</td>
</tr>
</tbody>
</table>

After establishing our main goals, we initially began by envisioning what a person would have to do, intend, and say, to demonstrate three qualitatively different levels for each epistemological belief dimension. We divided the instrument into three parts: the first part (MESSEB-Part A) would focus on participants’ behavior. However, because it is not possible to measure actual behavior in real world settings with a paper-and-pencil instrument, the first part is designed to support inferences based of epistemological judgments. That is, instead of actual behavior, we are settling on justifications and judgments made in the face of competing expert opinions, which fits into the third common method mentioned above that personal epistemological researchers employ to infer epistemological beliefs. The second part (MESSEB-Part B) is to measure intentions and the third part (MESSEB-Part C) is to capture explicit beliefs. We anticipated the three types of evidence for each epistemological belief dimension would converge at one of the three qualitative levels (i.e. dualism, relativism discovered, and contextual relativism), lending confidence to inferences.
MESSEB-Part A (Appendix A) measures epistemological judgments from problem scenarios involving competing assertions. Chosen problem scenarios have the quality of being open to multiple resolutions, which is the cornerstone of contextual relativism. Specifically, problem scenarios concern: a) potential introduction of a minimum-wage law in Hong Kong, b) benefits of breastfeeding verses bottle-feeding, and c) impact of media violence on children. We decided to have more than one scenario, because it is known that epistemological judgments can vary somewhat in different problems (see Briell et al., 2011; King & Kitchener, 2004). It is expected that three scenarios give us a better idea of the participant’s typical performance than a single scenario could. For each scenario, participants read opposing perspectives, answer probe questions, and give advice on a hypothetical problem. The first two probe questions are from Kuhn, Cheney, and Weinstock (2000). They are designed to support inferences about the Diversity of knowledge: “Is there more than one correct answer to (the respective issue)” An answer of “No” supports an inference of dualism. If participants answer “Yes”, they are asked “… are some answers better than others.” “No” supports a relativism discovered inference and “Yes” supports a contextual relativism inference. To understand participants’ epistemological judgments about the Stability of knowledge, the next question asks: “Will our understandings of (the respective issue) change over time?” The responses “No, hardly at all”, “Yes, drastically”, and “Yes, somewhat” support dualism, relativism discovered, and contextual relativism, respectively. Participants then give advice to fictitious individuals or entities that are in the midst of a decision involving competing assertions from scenario. Advice given is to be gleaned for epistemological judgments regarding Source of knowledge, Justification of knowledge, and Organization of knowledge dimensions.

MESSEB-Part B (Appendix B) aims to uncover intentions concerning the five relevant belief dimensions. By intentions, we mean beliefs consistent with each dimension that predispose the individual to certain actions. Participants are asked to assume the role of a teacher as they choose particular statements that would best inspire their teaching. We ask participants to imagine themselves as teachers, because teaching is particularly predisposed to the process of knowing. For each dimension a set of three choices is given – each choice supports dualism, relativism discovered, or contextual relativism. Options are designed with the rationale that if a participant were to endorse a particular choice as being inspirational to their teaching, it is because they hold a dimensional belief at the respective level. For instance, to be inspired by the statement, “Students must prove the absolute correctness of
their answers to their teachers’ questions,” suggests the participant believes knowledge must be proven to represent the objective reality, which is dualism for the *Justification of knowledge* (see above Table 1).

MESSEB-Part C (Appendix C) aims to determine what participants would say explicitly about each belief dimension. Because we anticipate it would be difficult for them to put the abstract epistemological beliefs into words (see Briell, Elen, Depaepe, & Clarebout, 2010), we sought alternative ideas for getting participants to explicitly express their beliefs. We settled on the idea of free response to stimulus statements, hoping that the statements would provide a helpful reference point for participants to share their beliefs. Each stimulus statement is intended to guide participants’ attention to the relevant epistemological belief dimension. Statements are worded from a theoretically naive perspective (i.e. dualism) to prevent simple recognition of a more sophisticated response (i.e. relativism discovered or contextual relativism). Participants choose “mostly agree” or “mostly disagree” with each naïve statement and explain why. It is expected participants would either accept the naïve statement (i.e., “mostly agree”) and explain why or choose “mostly disagree” and explain a more advanced view.

The design incorporates many of the “rules of thumb” endorsed by Bouchard (1976), including avoiding ambiguity, words with negative connotation, negatively worded sentences, and double negatives, keeping items as short as possible, and writing questions at a level of language appropriate for respondent population. The guide to instrument development by Frary (1996) was influential in the design of response selections.

The MESSEB was pretested on two Hong Kong college students who completed the survey and provided feedback. It was also sent to fellow researchers who were approached about piloting the instrument. They were queried about the face validity of instrument items and encouraged to provide comments and suggestions. The MESSEB was edited based on the advice of the fellow researchers queried and the data and comments from the two students who pretested it.

2) *Survey administration.* Participants were sent a copy of the MESSEB with attached cover letter and form for filling in relevant background information (namely, gender, age, class ranking, and nationality). They were asked to work alone, not to look up answers, and to
complete survey questions in the order presented. Completed surveys were returned to the first author via email.

**Data analyses**

For questions with response options, the chosen response was simply recorded as corresponding to dualism, relativism discovered, and contextual relativism. Open-ended questions in MESSEB-Part A and MESSEB-Part C required independent coding. In MESSEB-Part A, advice provided in three problem scenarios had to be coded for *Source of knowledge, Justification of knowledge* and *Organization of knowledge* dimensions. For MESSEB-Part C, explanations for choice of “mostly agree” or “mostly disagree” to naïve dimensional statements had to be coded for all five theoretical dimensions. In both cases: the first author examined these data and a developed coding scheme based on the theoretical model (i.e., one scheme MESSEB-Part A and one scheme for MESSEB-Part C). The first author and a second researcher then independently scored the data using the respective coding scheme. This resulted in 81 percent agreement for MESSEB-Part A scores (21 discrepancies of 108 epistemological judgment scores) and 85 percent for MESSEB-Part C scores (9 discrepancies of 60 explicit belief scores). Discrepancies were discussed with an agreement reached in each instance.

Coding determined a qualitative level (i.e., dualism, relativism discovered, or contextual relativism) for epistemological judgments in each problem scenario (i.e., three scores), for intentions, and for explicit beliefs per dimension. In some instances, the lack of data provided did not permit a definitive level. In some cases the data were sufficient for an approximation of relativism discovered to contextual relativism, otherwise no level was assigned (i.e., marked as insufficient data). Approximations and insufficient data occurred when epistemological judgment responses for MESSEB-Part A or explicit belief responses for MESSEB-Part C were insufficient.

We expected assigned epistemological judgment levels to fluctuate over the three problem scenarios (see Briell et al., 2011; King, Kitchener, & Wood, 1994). In concordance with that expectation, a prominent level (i.e., typical performance) was assigned based on the three individual-level assignments. When two or three individual levels concurred, we assigned this as the prominent level. When all epistemological judgment levels differed, we designated it as “no prominent level”.

Electronic Journal of Research in Educational Psychology, 11(2), 473-500. ISSN: 1696-2095. 2013, no. 30  
http://dx.doi.org/10.14204/ejrep.30.13035
Table 2. Individual Epistemological Judgments (IEJ) and Prominent Epistemological Judgments (PEJ) per Dimension

<table>
<thead>
<tr>
<th>Participant</th>
<th>Source</th>
<th>Justification</th>
<th>Organization</th>
<th>Diversity</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
</tr>
<tr>
<td>2</td>
<td>CR, D, D</td>
<td>NPL</td>
<td>D, CR, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
</tr>
<tr>
<td>3</td>
<td>ID, RD, ID</td>
<td>ID, ID, D</td>
<td>ID, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
</tr>
<tr>
<td>4</td>
<td>ID, ID</td>
<td>ID, D, D</td>
<td>ID, D, D</td>
<td>ID, D, D</td>
<td>ID, D, D</td>
</tr>
<tr>
<td>6</td>
<td>RD, D, D</td>
<td>D, CR, RD</td>
<td>CR, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
</tr>
<tr>
<td>9</td>
<td>RD, D, D</td>
<td>D, RD, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
</tr>
<tr>
<td>11</td>
<td>RD, CR RD</td>
<td>RD, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
<td>D, D, D</td>
</tr>
</tbody>
</table>

Note. D = Dualism. RD = Relativism Discovered. CR = Contextual Relativism. NPL = No Prominent Level. ID = Insufficient data.

Results

Table 2 above indicates individual and prominent theoretical level assignments per dimension. While we expected individual epistemological judgment scores to fluctuate, we anticipated the prominent epistemological judgment, the intention, and the explicit belief scores would concur at one of the three theoretical levels, lending confidence to inferences foreseen. However, as is evident in Table 3 below, this was not the case. Anticipated convergence at a certain level only occurred nine times out of the sixty sets. Less than half of the sets (i.e., twenty nine) concurred in two of the evidence types, suggesting only possible convergence at a certain level. These results indicate we are unable to draw comprehensive inferences about theoretical levels for the five epistemological belief dimensions as originally intended.
Table 3. Levels per Evidence Type per Dimension

<table>
<thead>
<tr>
<th>Participant</th>
<th>Source</th>
<th>Justification</th>
<th>Organization</th>
<th>Diversity</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PEJ</td>
<td>I</td>
<td>ExB</td>
<td>PEJ</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>NPL</td>
<td>CR</td>
<td>D</td>
<td>D</td>
<td>RD*</td>
</tr>
<tr>
<td>4</td>
<td>ID</td>
<td>CR</td>
<td>D</td>
<td>ID</td>
<td>RD*</td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>CR*</td>
<td>CR*</td>
<td>D*</td>
<td>CR</td>
</tr>
<tr>
<td>6</td>
<td>D*</td>
<td>D*</td>
<td>CR*</td>
<td>CR*</td>
<td>CR*</td>
</tr>
<tr>
<td>7</td>
<td>D</td>
<td>CR*</td>
<td>CR*</td>
<td>D*</td>
<td>CR</td>
</tr>
<tr>
<td>8</td>
<td>CR**</td>
<td>CR**</td>
<td>CR**</td>
<td>CR**</td>
<td>CR**</td>
</tr>
<tr>
<td>9</td>
<td>D*</td>
<td>CR</td>
<td>D*</td>
<td>D*</td>
<td>CR</td>
</tr>
<tr>
<td>10</td>
<td>CR*</td>
<td>CR*</td>
<td>D</td>
<td>ID</td>
<td>RD*</td>
</tr>
<tr>
<td>11</td>
<td>RD</td>
<td>CR</td>
<td>D</td>
<td>D*</td>
<td>D*</td>
</tr>
<tr>
<td>12</td>
<td>NPL</td>
<td>CR*</td>
<td>CR*</td>
<td>D*</td>
<td>CR</td>
</tr>
</tbody>
</table>

Our next research questions concerned dimension performance and combined dimension performance. Because we could not base dimension performance on assigned qualitative levels, we determined a *dimensional average* for each epistemological belief dimension. For each evidence type (i.e., prominent epistemological judgment, intention, and explicit belief), a score of 1, 2, or 3 was given for dualism, relativism discovered, or contextual relativism, respectively. (If the prominent epistemological judgment level is “no prominent level,” we gave a score of 2.) Evidence type scores were subsequently added together and divided by three. (If one or more evidence type scores were marked “insufficient data,” we did not calculate a dimensional average for the respective dimension.) Dimensional averages were then averaged together (excluding instances of insufficient data) to determine a *combined dimension average*. Dimensional averages and combined dimension averages can be found in Table 4 below.

Participants as a group scored lowest for the dimensions of *Justification of knowledge* (M = 1.9, SD = .54) and *Organization of knowledge* (M = 1.8, SD = .58), meaning many participants showed evidence of dualism to relativism discovered. Of note, *Organization of knowledge* also had the most instances of insufficient data. Participants scored highest for *Stability of knowledge* (M = 2.7, SD = .32), indicating many participants showed evidence of contextual relativism in this dimension. The combined dimension average for participants as a group is a moderate score (M = 2.3, SD = .40). Of relevance, half of the combined dimension averages had to be calculated excluding insufficient data. The lowest combined dimension average for a single participant is 1.5 and the highest is 2.9, indicating highly dualistic scores and highly contextual relativistic scores, respectively, in all dimensions.
Table 4. Dimensional Average and Combined Dimension Average (CDA)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Dimensional Average</th>
<th>CDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Justification</td>
<td>Organization</td>
</tr>
<tr>
<td>1</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>3</td>
<td>ID</td>
<td>1.3</td>
</tr>
<tr>
<td>4</td>
<td>ID</td>
<td>ID</td>
</tr>
<tr>
<td>5</td>
<td>2.3</td>
<td>1.7</td>
</tr>
<tr>
<td>6</td>
<td>1.7</td>
<td>2.3</td>
</tr>
<tr>
<td>7</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>8</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>9</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>10</td>
<td>2.3</td>
<td>ID</td>
</tr>
<tr>
<td>11</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>12</td>
<td>2.7</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Group 2.2* (SD = .39) 1.9* (SD = .54) 1.8* (SD = .58) 2.2 (SD = .46) 2.7 (SD = .32) 2.3 (SD = .40)

Note. Dimensional Averages = (Prominent Epistemological Judgment + Intention + Explicit Belief)/3. Dimensional Average of 1 would indicate dualism scores on all evidence type data while a score of 3 would indicate contextual relativism scores for all evidence type data. *Insufficient data excluded in calculation.

Discussion

We anticipated participants would be at a consistent theoretical level across evidence types for each dimension. If this were the case, we would have felt justified inferring theoretical levels for epistemological belief dimensions. However, it is obvious from Table 3 that participants are seldom at consistent levels of theoretical sophistication for the three types of evidence collected. This means the data do not permit self-evident inferences of theoretical levels per dimension.

As a group, participants performed worst in the Organization of knowledge and Justification of knowledge dimensions, meaning our small sample did not have a solid grasp that knowledge is highly connected in the mind and did not understand the importance of evidence and reasons to support one’s knowledge. Considering the average age of participants and the range of education levels (i.e., entry-level to doctoral-level university students), somewhat higher scores were expected. The findings of Baxter Magolda (1992, 2004), Perry (1970, 1981), and Schommer (1990) all suggest students’ beliefs about the nature of knowledge tend to advance with age and schooling. As a group, they had good understanding that human knowledge progresses over time (Stability of knowledge). This result parallels a
finding by Chan and Elliott (2004) that Hong Kong teacher education students tended to believe that knowledge changes over time. We determined participants as a group had a moderate combined dimension average (i.e., in the range of relativism discovered to contextual relativism).

Significantly, there were ten instances for which a dimensional average could not be determined due to missing or inadequate data in one or more of the evidence types. This also meant half of the combined dimension averages were calculated excluding these scores. Modifications to the instrument might remedy these difficulties (see below).

Conclusions

We set out to determine if it were possible to create a paper-and-pencil instrument that permits adequate inference of participants’ highly complex epistemological beliefs in five theoretically relevant dimensions for three qualitatively different levels. Our testing of the MESSEB demonstrates the challenge of doing so.

Our data suggest participants vary in theoretical level according to the type of evidence collected. One explanation for the variability is that the MESSEB is not correctly eliciting the intended evidence type data. Clearly, further testing and amendments are requisite to ensure instrument items are valid and reliable. Potential amendments are discussed below. A second explanation is that the MESSEB is eliciting the intended evidence type data, but these evidence types are not appropriate for inferring the intended general epistemological belief dimensions. If this is the case, then three vantages of the same construct have not been measured, but of more than one construct. The question then turns to what other kinds of evidence would better enable proper inference of the general epistemological belief dimensions sought. The MESSEB is designed in consideration of Pajares (1992), but other attempts and measurement theory must also be explored.

A third explanation is that although the same construct is being measured, variation is genuine and typical. There are a couple of reasons for why different types of evidence measuring the same construct might vary. One reason is that individuals do not have epistemological beliefs – at least not in the sense that they are firmly held and bound by coherent logic. That is, predominant understanding of the construct is flawed. Along this line
of thinking, is the suggestion that epistemological beliefs are not robust in various contexts. Instead, individuals have a host of beliefs (referred to as epistemological resources by the theorists) that are sensitive to and specifically activated in certain contexts (see Hammer & Elby, 2002; Louca, Elby, Hammer, & Kagey, 2004). If the theory by Hammer and colleagues is correct, it is possible that our questions and prompts led participants to think of certain contexts when they responded – as opposed to the context-general beliefs we were seeking. However, the acceptance of this theory diminishes the relevance of epistemological beliefs. By presuming that epistemological beliefs exist as a host of epistemological resources that fluctuate with and are tightly bound to the context than the context is pertinent and the beliefs are of secondary importance, if not inconsequential. An alternative reason is that inconsistencies are completely “stable and harmonious” within a network of beliefs at multiple levels of priority. Pajares (1992) provides the illustration of a teacher who loved his students and was committed to their welfare, yet he had a stern and authoritarian demeanor towards them. When understood in full, this teacher’s actions were not inconsistent with his beliefs: “His strict attitude was based on the belief that the development of a young man’s character required toughness and discipline, and this was of greater concern to him than his desire to show the tenderness and affection he felt” (p. 319). Hence, incongruent evidence may be perfectly coherent within the person’s entire network of beliefs. If the third reason is correct, then the more thorough the investigation the better understood the beliefs would be. In the case of our study, perhaps observations of actual behavior in conjunction with follow-up interviews would have revealed the convergent evidence sought and shed light on seeming inconsistencies.

Either the second or third explanation for variability, if accurate, would have far reaching implications for the way personal epistemological researchers infer epistemological beliefs. Since researchers normally examine only one evidence type, it is clear that the inferences gleaned would likely be different if other types of evidence were considered. This, of course, puts in question even further the value of current instrumentation, suggests different types of instruments would routinely give different results, and poignantly demonstrates the complexity of epistemological beliefs and the difficulty of drawing inferences about them.

It is clear modifications to the MESSEB are warranted. Above all, additional attempts could be made to verify the reliability and validity of items. In conjunction, it is especially important to reduce instances of non-response and inadequate response. When participants
were asked about their overall impression of the instrument, they uniformly responded that it was time consuming (over 1 hour) and required too much writing. Reducing the instrument length and providing more response selections may mitigate these objections, simultaneously making the instrument easier to score. To reduce the instrument length, one of the three problem scenarios in the epistemological judgment section (i.e., MESSEB-Part A) could be eliminated. Maintaining at least two scenarios is desirable to provide some gauge of typical or average performance. It might also be possible to incorporate response selections in MESSEB-Part C that allow participants to explicitly endorse certain beliefs, as opposed to explicitly explaining them. A statement constituting an explicit belief for each dimension could be followed by a set of reactions to choose from. Each choice within the set could represent endorsement of an explicit belief for one of the three qualitative levels. For instance, to the dualistic statement for the Source of knowledge (i.e., “Learning is best described as a direct transfer of knowledge from those who have it to learners who receive it [e.g., parents give their knowledge to their children]”), the response selection “Disagree - because the learner is also very important in the learning process” would be consistent with Contextual relativism. The drawback to response selections is that it may allow participants with less sophisticated beliefs to simply recognize a more sophisticated one (see King & Kitchener, 2004). It may help alleviate the aforementioned drawback if the dimensional statements are written at random levels of theoretical sophistication; though this still does not prevent recognition of a sophisticated reaction to a naïve dimensions statement.

A final modification might also be considered. It could be argued that MESSEB-Part B prompts participants to consider only formal knowledge. Currently the participant is asked to respond from the point of view of a teacher instructing students. While the profession of teaching is intrinsically associated with the process of knowing, it may have the effect of drawing participants’ attention away from informal knowledge. Hence, it may be necessary to reconstruct this part to elicit a broader understanding of knowledge in general.

To move forward in the many significant questions raised in this pilot, it would be beneficial to retest a revised MESSEB once modifications are made. The additional testing may shed light on (a) whether variation in the different evidence types persists and (b) the value of the evidence collected. Regarding the former, the question is whether irregular evidence type performance is the norm or the exception. Regarding the latter, it would be possible to examine potential relationships between dimensional data and dependent variables.
of theoretical importance (e.g., academic achievement, task performance). Such research may provide a better idea of whether MESSEB data are empirically relevant to indicators of academic success. If it is determined that irregular performance is the norm and dimensional data are not significant to indicators of academic success as theoretically supposed, then a shift in approaches would be justified. Namely, a ground-up study of epistemological beliefs that thoroughly considers the construct from as many vantages as possible might be instrumental to explain variation, identify others kinds of data that more effectively reveal the construct, and enable meaningful refinements to our theoretical model and ultimately to the MESSEB.

Acknowledgements

Many thanks to Nick Stassens for his significant efforts in data coding. We would also like to thank the reviewers for their constructive comments and advice.
References


Bouchard, T. J. (1976). Field research methods: Interviewing, Questionnaires, Participant observations, systematic observations, unobtrusive measures. In M. D. Dunnette (Ed.) Handbook of industrial and organization psychology (pp. 363-413). Chicago: Rand McNally.


Appendix A: MESSEB – Part A (Epistemological Judgments)

Directions: Read the opposing arguments for each topic before answering the questions that follow.

Reading Topic 1: Should Hong Kong Introduce a Minimum-Wage Law?
- [Argument 1: Pro-minimum wage argument]
- [Argument 2: Anti-minimum wage argument]

Reading Topic 2: Breastfeeding vs. Bottle-feeding
- [Argument 1: Pro-bottle-feeding argument]
- [Argument 2: Pro-breastfeeding argument]

Reading Topic 3: The Impact of Media Violence on Children
- [Argument 1: Entertainment violence can have negative impact on children]
- [Argument 2: Entertainment violence does not have negative impact on children]

Is there more than one answer to [Topic]? (Diversity of knowledge)
- Yes
- No

If yes to the above question, can some answers be better than others? (Diversity of knowledge continued)
- Yes
- No

Do you suppose our understandings of [Topic] will change over time? (Stability of knowledge)
- Yes, drastically
- No, hardly
- Yes, somewhat

Questions for Source of knowledge, Justification of knowledge, Organization of knowledge:
- (Topic 1) Hong Kong lawmakers are considering instituting a minimum-wage law, which could have an important impact on the lives of Hong Kong citizens. Currently, Hong Kong has no legal minimum wage. What advice would you give Hong Kong lawmakers concerning the introduction of a minimum-wage law? Please explain.

- (Topic 2) Kate is an expecting mother who works as a flight attendant for a major airline. Should Kate breastfeed her baby? Please explain.

- (Topic 3) Norton is a media producer for a major television network. His network has bought the rights to show a number of popular movies throughout the summer season. The network has received frequent complaints from concerned parents about the violent content of these films. How should Norton reply to the parents? Please explain.
Appendix B: MESSEB – Part B (Intentions)

Directions: Imagine you are a teacher as you answer the questions below. For each question, only circle the one statement that would best inspire your teaching.

Which of the three statements would best inspire your teaching? (Source of knowledge)

1. Teachers are fully responsible for what their students learn.
2. Both teachers and students are responsible for what is learned.
3. Students are fully responsible for their own learning.

Which of the three statements would best inspire your teaching? (Diversity of knowledge)

1. When a student’s answer is different from the teachers, the student should receive high marks if the answer is reasonable.
2. When a student’s answer is different from the teachers, the student should be praised for independent thinking even when the answer is unreasonable.
3. When a student’s answer is different from the teachers, the student should be guided to the correct answer.

Which of the three statements would best inspire your teaching? (Organization of knowledge)

1. Music should be taught in music class, reading should be taught in reading class, math should be taught in mathematics class, etc.
2. It is not important whether the teacher interconnects learning material for different courses in a lesson.
3. Language can be taught with music, math with a cooking demonstration, science at a field trip to an art exhibition, etc.

Which of the three statements would best inspire your teaching? (Stability of knowledge)

1. Students today must learn essentially the same knowledge as their parents had when they were young.
2. Students today need to learn knowledge that is somewhat different than their parents had when they were young.
3. Students today need to learn knowledge that is completely different than their parents had when they were young.

Which of the three statements would best inspire your teaching? (Justification of knowledge)

1. Students should feel their understandings of the subject matter are respected, regardless of what those understandings are or how they came about.
2. Students must prove the absolute correctness of their understandings about subject matter.
3. Students should be expected to provide reasonable, well-defended answers to their teachers’ questions about subject matter.

Appendix C: MESSEB – Part C (Explicit Beliefs)
Directions: Below are five statements about the nature of knowledge and knowing. Please choose whether you mostly agree or disagree with each and explain why.

Learning is best described as a direct transfer of knowledge from those who have it to learners who receive it (e.g., parents give their knowledge to their children). (Source of knowledge)

- Mostly agree
- Mostly disagree

Please explain.

Every problem has a single correct answer that can be known with certainty. (Diversity of knowledge)

- Mostly agree
- Mostly disagree

Please explain.

The many things a person knows about the world can be found in his/her mind as unconnected pieces information that have been categorized according to subject area. (Organization of knowledge)

- Mostly agree
- Mostly disagree

Please explain.

Knowledge does not change much from generation to generation. (Stability of knowledge)

- Mostly agree
- Mostly disagree

Please explain.

A person is justified in saying, “I know,” only when his/her beliefs are proven to correspond to the objective reality. (Justification of knowledge)

- Mostly agree
- Mostly disagree

Please explain.