

Assessing Levels of Epistemological Understanding: The Standardized Epistemological Understanding Assessment (SEUA)

Natalia Żyluk¹ · Karolina Karpe¹ · Mikołaj Michta¹ · Weronika Potok¹ · Katarzyna Paluszkiewicz¹ · Mariusz Urbański¹

Published online: 18 April 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract This article describes the process of modification and Polish adaptation of an instrument constructed to assess the level of epistemological understanding. The original tool was developed by Kuhn et al. (Cogn Dev 15:309–328, 2000) in order to account for transitions between, and coordination of, subjective and objective dimensions of knowing across different judgement domains (the domains of personal taste, aesthetics, moral values, truths about the social world and truths about the physical world). Our aim was to improve its psychometric properties. The main changes included extending the list of test items, a new administration procedure and the introduction of a quantitative scoring method. The outcome is a valid, reliable and standardised instrument—the Standardized Epistemological Understanding Assessment.

Keywords Epistemology · Epistemological understanding · Personal epistemology · Epistemological development · Judgement domains

1 Introduction

The aim of this paper is to present the process of modification and Polish adaptation of a test instrument as described in *The development of epistemological understanding* by Kuhn et al. (2000). It was meant to be used in

assessing the level of epistemological understanding in five judgement domains: personal taste, aesthetics, moral values, truths about the social world and truths about the physical world. The instrument developed by Kuhn et al. has been used, for example, in research on the role of epistemic thinking in online learning processes (Barzilai and Zohar 2009); sociocultural determinants of epistemological understanding (Tabak and Weinstock 2008); the influence of epistemological views and interest in the interpretation of controversial text and topic-specific belief changes (Mason and Boscolo 2004) as well as the relationship between gender, grades and curriculum, and the domain-dependent level of epistemological understanding (Mason et al. 2006). In some of the studies, significant changes to the procedure were suggested. Ahola (2009) asked participants to provide justification of their judgements in some domains, proving them inconsistent in some cases. Mason et al. (2006) interviewed a limited number of the participants for clarification of their given answers in certain domains (but not for individual test items) after the tool was administered. Christodoulou et al. (2010) as well as Mason and Boscolo (2004) introduced alternative scoring methods.

Drawing on this previous research we suggest some further modifications to the original instrument, in particular an extension of the number of test items, in order to improve its psychometric properties. The outcome is a valid, reliable and standardized instrument—the Standardized Epistemological Understanding Assessment (SEUA).

This paper is structured as follows. In Sect. 2 we outline the theoretical assumptions, on which the original instrument is based. In Sect. 3 we describe the details of our study. In Sect. 4 the results are given, followed by discussion and plans for future research in Sect. 5.

✉ Mariusz Urbański
murbansk@gmail.com

Natalia Żyluk
nzyluk@gmail.com

¹ Institute of Psychology, Adam Mickiewicz University in Poznań, Szamarzewskiego 89 AB, 60-568 Poznań, Poland

2 Theoretical Assumptions

Investigations concerning the formation of beliefs as to the nature of knowledge and knowing are an important line of inquiry within educational studies and developmental psychology. Perry (1970) is considered to be the person who initiated empirical research on this subject, but, as Hofer and Pintrich (1997) stressed, the origin of personal epistemology can be associated with Piaget's theories of cognitive development (1950). Since then, a few distinct lines of research on this topic have emerged and progressed. We shall follow the line according to which the construct of personal epistemology is construed as a processual and developmental one rather than as a relatively static system of beliefs (Ahola 2009). Proponents of this approach distinguished between consecutive stages at which an individual's attitude towards the nature of knowledge and knowing can be characterised; each of these stages is substantially different from the others (see e.g. Baxter Magolda 2004; Kitchener and King 1981; Kuhn 1999; Perry 1970). In this paper, we characterize the process of the adaptation and modification of the instrument based on Kuhn's model of cognitive development (Kuhn 1999, 2000) in which coordination of the subjective and objective aspects of knowing plays the main role in the epistemological progression of an individual.

2.1 Levels of Epistemological Understanding

As Kuhn et al. (2000) note, the cognitive and intellectual functioning of an individual is significantly determined by his or her views on what knowledge is and how it is evaluated and acquired. These individual conceptions of knowledge and knowing determine one's level of epistemological understanding. As we just mentioned, according to the authors we refer to, changes in the relationship between subjective and objective dimensions are responsible for views on the nature of knowledge and belief: formation of a mature epistemological understanding is a process that starts with radical objectivism (knowledge as a certain and objective entity), leading through subjectivism, to an integration of both dimensions (allowing for uncertainty and the possibility to evaluate beliefs) (Kuhn 1999). The way that personal epistemology develops was characterized with its emphasis on different aspects and dimensions (see the review by Hofer and Pintrich 1997), but in many cases the general schema of such changes is similar to that described above.

Kuhn et al. (2000) distinguished between four levels of epistemological understanding: realist, absolutist, multiplist and evaluativist. Assessment of the realist understanding, as being typical only for early childhood, was not

included in their instrument. Realist and absolutist see knowledge as an objective entity, completely knowable and intellectually accessible. In the absolutist view, knowledge is certain and refers to a reality external to the subject. The difference between a realist and an absolutist can be seen in their approach to assertions: for a realist assertions are copies of objective reality, while an absolutist treats assertions as facts that represent objective reality in a correct or incorrect way. Under the absolutist interpretation, when two people come to the different conclusions it cannot be the case that both of them are right, since there can be only one, "ultimate", reality they can refer to. The absolutist allows for the possibility of false belief.

In order to transfer from the absolutist to the multiplist level one must realize the uncertainty of knowledge and its subjective side. For a multiplist, knowledge has multiple sources and is seen as closely tied to the perceiving subject, therefore in this view the objective dimension is simply abandoned. Individuals on a multiplist level perceive all judgements as merely opinions, and as everybody has a right to have one, in their view all judgements—even conflicting ones—can be equally right. The view that knowledge is a product of the human mind rather than an externally located entity, is the reason why it is regarded as uncertain.

At the evaluativists' level an integration of the objective and subjective sides of knowing occurs. For the evaluativist, as for the multiplist, knowledge is uncertain and considered to be constructed by people, but—and this reflects the objective aspect of this epistemic level—in assessing different views the evaluativist takes into account the empirical evidence or support of persuasive argumentation. The evaluativist allows for the simultaneous rightness of two incompatible judgements, but prefers the one that has more merit or is better justified.

2.2 Epistemological Levels and Judgement Domains

Analyses of developmental changes in epistemological understanding give rise to the question of whether transitions from one level to another are somehow domain-dependent. Kuhn et al. considered this problem within the context of the following domains: personal taste judgements, aesthetic judgements, value judgements and truth judgements. They further differentiated truth judgements into two categories: truth judgements in a social context and truths within the context of the physical world.

The authors expected the transition from an absolutist to a multiplist level to occur first in the judgement domain of personal taste, then in aesthetic judgements, later in the domain of values, followed by the domain of truths in a social context and finally within the domain of truths about the physical world. They also suggested that the transition from

the multiplist to the evaluativist level may occur in the reverse order. Within the domain of personal taste, the transition from the multiplist to the evaluativist level may not happen at all.

2.3 The Original Instrument

To verify their hypotheses Kuhn et al. (2000) developed a test instrument to assess the level of epistemological understanding, which can be used to determine if a transition from one level to another has taken place.

This assessment instrument consisted of 15 pairs of sentences—3 for each judgement domain. Each pair consisted of two mutually incoherent judgements, presented by two people: Chris and Robin. Sample pairs for each judgement domain are:

Judgements of personal taste:

Chris says cool autumn days are nicest.
Robin says warm summer days are nicest.

Aesthetic judgements:

Robin thinks the first painting they look at is better.
Chris thinks the second painting they look at is better.

Value judgements:

Robin thinks lying is wrong.
Chris thinks lying is permissible in certain situations.

Judgements of truth about the social world:

Robin has one view of why criminals keep going back to crime.
Chris has a different view of why criminals keep going back to crime.

Judgements of truth about the physical world:

Robin believes one book's explanation of what atoms are made up of.
Chris believes another book's explanation of what atoms are made up of.

To assess if the transition from the absolutist level to the multiplist level has occurred in an individual, for each pair of sentences the question posed is "Can only one of their views be right, or could both have some rightness?". The diagnostic answer for the absolutist level is "Only one view can be right". If one answers "Both could have some rightness", the following question is asked in order to assess the transition from the multiplist to the evaluativist level: "Could one view be better or more right than the other?". The answer "One could not be more right than the other" is the diagnostic answer for the multiplist level, while "One could be more right" is the diagnostic answer for the evaluativist level.

In the original tool, the instruction for marking answers is as follows:

Can only one of their views be right, or could both have some rightness?

ONLY ONE RIGHT

BOTH COULD HAVE SOME RIGHTNESS (circle one)

IF BOTH COULD BE RIGHT:

Could one view be better or more right than the other?

ONE COULD BE MORE RIGHT

ONE COULD NOT BE MORE RIGHT THAN THE OTHER (circle one)

An individual is assigned a category (an absolutist—A, a multiplist—M, or an evaluativist—E) within a given judgement domain, if for at least two out of three statements in a domain, the questions are answered in a way characteristic for one of those levels of epistemological understanding. When an individual responded in three different ways within one judgement domain, then he or she is assigned the category M within that domain (see Table 1).

Given the answers, every participant can be assigned a specific profile, e.g. MMAEE, where consecutive letters represent levels of epistemological understanding in the domains of personal taste, aesthetics, value judgements, truth judgements about the social world and truth judgements about the physical world, respectively.

In the study described by Kuhn et al. (2000), participants took the test in small groups. Bearing in mind the results of studies that suggest that views on the nature of knowledge and beliefs of an individual can be influenced by her education, life experience and age (see Hofer and Pintrich 1997, as cited by Kuhn et al. 2000), the group included adults that varied with respect to these characteristics as well as children from middle childhood to adolescence. The instrument was administered in paper-and-pencil form. Pairs of statements were set in random order. Participants were accompanied by a researcher, who could have been

Table 1 Assignments of the basic categories (original instrument)

A	M	E	Category
3	0	0	A
2	1	0	A
2	0	1	A
0	3	0	M
1	2	0	M
0	2	1	M
1	1	1	M
0	0	3	E
1	0	2	E
0	1	2	E

asked (explanatory) questions. Taking the test took 10–20 min.

2.4 Instrument's Original Limitations

The instrument presented by Kuhn et al. (2000) was considered a preliminary attempt to create a test tool for levels of epistemological understanding. We analysed it in order to identify aspects that could benefit from further improvements. The content diversity of the test items in particular domains is relatively small. This could have led to participants connecting the statements easily in larger clusters and answering all the elements in a cluster in a similar way without proper reflection. What is more, answers regarding some pairs of statements could have been influenced by their subject matter. Therefore having only three pairs in each category carry a high risk of switching the participant's score from one style to another by answering questions concerning only one pair which is not properly formulated. The authors did not provide a quantitative scoring method that could be easier to use than the profiles of epistemological understanding in the domains.

2.5 SEUA Version of the Tool: The Main Changes

Our work on Kuhn's et al. (2000) instrument consisted of: translating original pairs of sentences into Polish, extending the original list of sentences with new test items, changing the administration procedure (from a paper-and-pencil version to that of a recorded interview), changing the instructions and materials for both the experimenter and the subject and introducing a quantitative scoring method (converting the nominal values to numbers).

The translation process consisted of a few phases. First, four English to Polish translation versions were proposed (developed independently), one of which was selected by Polish native speakers with advanced English levels as the one that most appropriately reflected the meaning of the original sentences. Subsequently, the chosen versions were sent to a proficient speaker of both English and Polish in order to check and add necessary corrections.

Ten test items were added to the original version of the instrument (two in each domain), chosen by means of the competent judges method. It should be noted that, as a result, the original version of the instrument is nested within our extended version. One benefit from extending the number of pairs in each domain from three to five is to lower the possibility of the overall score (and thus an epistemological level of understanding ascribed to an individual in a domain) being influenced by a particular topic of a certain pair of statements. One can easily imagine a situation, in which an incorrect understanding of one pair of sentences during the test can occur: the participant can have very

strong feelings regarding the topic, can be distracted for a moment or mishear the sentences. In such cases, the given answer may not be an appropriate indicator of the level of epistemological understanding. When a distorted answer of this type is 1/3 of the total score in a given domain, it has a larger influence on the overall score than in a situation where there are five test items in each domain.

For the remainder of this paper we shall use the following convention for naming the versions of the instrument: EUA for the original version, nEUA for the original version nested within the extended one, and SEUA for our fully extended version. Let us stress again, that EUA and nEUA contain the same test items but nEUA is administered as a part of SEUA.

Changing the test procedure was aimed at obtaining more accurate test results and getting valuable feedback about test items from the participants. It is worth mentioning, that the interview method was previously used in studies on personal epistemology, for instance by Kitchener and King (1981) in their analysis of reflective judgement. As we mentioned in the introduction, a few research studies that also employ EUA have introduced some changes in the original procedure, including interviewing elements (e.g. Mason et al. 2006; Ahola 2009).

The instructions were constructed to prevent the need of social approval influencing the answers. In particular, it was stressed that there are no right or wrong answers to the questions asked by the researcher and, due to individual differences, whether or to what extent a person allows for the simultaneous rightness of certain views can vary from person to person. Given that the study was conducted in the form of an interview, it was necessary to create conditions in which participants wanted to respond in an accurate and sincere manner.

Additional materials developed for this research included an answer card (for the experimenters to check the answers), a schema describing the answering procedure (for the participants), a list of test items (for the experimenters) and separate cards for each test item (for the participants). All the materials used in this research are available online at <http://reasoning.edu.pl/> (section: Research projects).

Finally, the development of a new scoring method enabled more thorough statistical analyses to be carried out (mainly to evaluate the various psychometric properties of the instrument in both the EUA and SEUA versions).

3 Method

3.1 Participants

The original study conducted by Kuhn et al. (2000) included seven groups of participants varying in age, life

and educational experience, as the main objective of the study was to assess if epistemological understanding develops in the predicted order across judgement domains. Due to the fact that the aim of our study was slightly different (the adaptation of the instrument with a modified procedure and evaluation of its psychometric properties), the group of participants was more homogeneous. The sample consisted of 40 adults with ages ranging from 19 to 35 ($M = 23.13$; $SD = 2.98$). The gender proportion was balanced, with 23 females and 17 males ($\chi^2 = 0.90$; $p > 0.05$); the sociodemographic data are presented in Table 6. For their participation in the procedure (two-step, see: Sect. 3.2), the subjects received gift cards (50 PLN) to a bookstore chain. All participants gave their informed consent before taking part in the experiment, in particular with respect to audio recording the sessions.

3.2 Procedure

The testing procedure consisted of two phases, separated by at least six-day intervals (maximum 21 days). During the first session participants were interviewed with the EUA version of the instrument (15 test items) and during the second one—with the SEUA version (25 test items), in which the nEUA is nested. During each of the sessions we tested the same group of participants. All interviews were conducted in the laboratory of the Reasoning Research Group at the Institute of Psychology, Adam Mickiewicz University in Poznań.

At the beginning of each session the participants were acquainted with the instructions and received response schema (presenting which questions should be answered and in what order for each pair of statements).

Subjects were presented with pairs of statements in written form, each pair on a separate card. An experimenter read these statements aloud and then gave a test card to a participant, so he or she could think about the answers. The participant answered the questions and gave explanations, in the order presented on the response schema. The schema visually presented the question order as in the original study, but also included indications that the participants should provide an explanation for each answer immediately after giving it.

As in the original paper-and-pencil procedure, in the case of each pair of statements participants were asked “Can only one of their views be right, or could both have some rightness?” The possible responses were “Only one can be right” or “Both could have some rightness”; in both cases the participants were asked for an explanation of the answer they chose. When a participant’s answer was “Only one can be right” then he or she was assigned the category A (for “absolutist”) for that pair of statements, and the experimenter read another pair. When the

participant gave the response indicating that both statements could have some rightness (and explained the reasons), a follow-up question was asked “Could one view be better or more right than the other?” In this situation the participant could react with “One could not be more right than the other” (category M for “multiplist” was assigned for that pair) or “One could be right” (category E was assigned for “evaluativist” for that pair). After providing an explanation for the chosen answer, the participant was given another pair of statements. The whole procedure was repeated 15 or 25 times depending on the version of the instrument.

For every pair of statements, the experimenter wrote down the answer given using a letter code for the three categories (A, M or E) on the answer card.

3.3 Scoring

3.3.1 Qualitative Scoring

In the original instrument, EUA, as well as in nEUA, a participant may be assigned one of the following three categories: A, M, or E in each of the judgement domains, depending on the number of answers that fall into a certain category (see Table 1). The scoring method for the original instrument was fully described in Sect. 2.3.

In the case of the SEUA version of the instrument we calculated the profiles both for the nEUA version (scored as described above) and for the SEUA version (scored in different way, as described below). As a result, every participant ended up with three profiles, obtained using EUA (in the first phase of the study) and SEUA with nEUA nested (in the second phase).

The profiles for the SEUA version were determined in a slightly different manner than for EUA and nEUA, as the domination of 3 out of 5 items was considered too weak to justify the assignment of a certain level and a distribution of 2–2–1 across categories was confusing. In addition to the original three letters we used the signs “+” and “–”, so for SEUA the possible categories were: A, A+, M–, M, M+, E–, E. To give an example: when an individual received “A+” category in a certain judgement domain it meant that he or she is an absolutist with a multiplist tendency in this domain, receiving a “M+” in a specific domain meant that he or she was a multiplist towards evaluativist, etc. Table 2 gives the details of the assignments of those additional categories. In order to receive a “clear” A, M or E category in SEUA with no additional signs, it was necessary to get at least 4 answers that fell into a certain category.

Our modifications are somewhat similar to the ones proposed by Christodoulou et al. (2010), who in EUA scoring also introduced categories such as A+ (“a mix of

Table 2 Assignments of the additional categories (SEUA only)

A	M	E	Category
3	2		A+
3		2	M-
	3	2	M+
2	3		M-
	2	3	E-
2		3	M+
2	2	1	M-
2	1	2	M
1	2	2	M+
3	1	1	A+
1	1	3	E-
1	3	1	M

absolutist and multiplist responses”), and E- (“2 evaluator responses and another response, which was either absolutist or multiplist”). They also used the “indeterminate” category, which was assigned to a participant who gave one response of each type.

3.3.2 Quantitative Scoring

Using quantitative scores makes it possible to assess the internal consistency of each sub-scale of the instruments (that is, test items concerning each particular domain), to check if scores are stable over time and to detect some of the instruments’ weaker points (e.g. pairs of statements that are negatively correlated with the rest of test items within one domain).

Besides a qualitative profile, participants were assigned points for every given answer, which were summed up within each domain and for the instrument as a whole. Subjects got scores for every domain and summary results separately for the EUA, nEUA and SEUA versions.

For every single answer A, 1 point was provided; for M—2 points; and for E—3 points. For each domain in EUA and nEUA it was possible to obtain from 3 to 9 points; the maximum summary score for the whole tool therefore equals 45. For SEUA the scores within domains ranged from 5 to 15, and for the whole tool participants can score a maximum of 75 points. Mason and Boscolo (2004) and Mason and Scirica (2006) also introduced summary scores for EUA which were interpreted as an indication of the general level of epistemological understanding. We consider summary scores as offering some information about how advanced the development of epistemological understanding is in an individual. However, due to the domain dependency of epistemological understanding we are not convinced that those scores can be interpreted as a reliable indication of its general level.

Table 3 presents scores that one can obtain depending on responses within one judgement domain (three test items) in EUA and nEUA. As the scores 5 and 7 can be received when two different categories are obtained (5: A or M and 7: M or E) it is necessary to supplement every score with the qualitative information of the category that a participant obtained.

Table 4 presents possible scores within one judgement domain (five test items) in SEUA. As for EUA and nEUA, for the results of SEUA to be informative they should be reported as a pair of category-score.

4 Results

All the statistical analyses were carried out using the statistical software SPSS v. 22.

4.1 Descriptive Statistics

Table 5 contains descriptive statistics for participants and test results for all the versions of the instrument and all the considered domains. The maximum and minimum for each version reflect the actual scores obtained by the participants; theoretically the highest number of obtainable points possible is 45 for EUA and nEUA, and 75 for SEUA. Table 6 contains the sociodemographic characteristics of participants.

4.2 Reliability

Internal consistency was assessed using Cronbach’s α coefficient for each judgement domain. The Wilcoxon signed-rank test was used to determine whether the differences in scores between measures were significant. Additionally, the Friedman test was conducted in order to compare the differences in participants’ profiles obtained using all three versions of the instrument. Both the Wilcoxon and the Friedman tests were considered to be forms of stability measures.

Table 3 Possible scores within one domain in EUA and nEUA

A	M	E	Category	Score
3			A	3
	3		M	6
		3	E	9
2	1		A	4
2		1	A	5
1	2		M	5
1	1	1	M	6
	2	1	M	7
1		2	E	7
	1	2	E	8

Table 4 Scores possible to obtain in the SEUA version

A	M	E	Category	Score
5			A	5
	5		M	10
		5	E	15
4	1		A	6
4		1	A	7
	4	1	M	11
1	4		M	9
	1	4	E	14
1		4	E	13
3	2		A+	7
3		2	M-	9
	3	2	M+	12
2	3		M-	8
	2	3	E-	13
2		3	M+	11
2	2	1	M-	9
2	1	2	M	10
1	2	2	M+	11
3	1	1	A+	8
1	1	3	E-	12
1	3	1	M	10

Table 6 Sociodemographic data

	N	%
Sex		
Male	17	42.50
Female	23	57.50
Studying		
Yes	31	77.50
No	9	22.50
Graduate		
Yes	28	70.00
No	12	30.00
Working		
Yes	21	52.50
No	19	47.50

4.2.1 Internal Consistency

The values of Cronbach’s α were calculated for each domain, for each version of the instrument. As none of them is homogenous, α was not calculated for the whole tests. In almost each domain, the highest α values are observed for the longer version (see Table 7). A significant increase in the reliability in most domains in the SEUA version, in comparison with the EUA, suggests that

Table 5 Descriptive statistics

	Min	Max	M	Me	SD	Kolmogorov–Smirnov test	
						Z	p
EUA—personal taste	5	8	6.40	6.00	0.67	0.374	<0.001
EUA—aesthetic	6	9	6.43	6.00	0.87	0.437	<0.001
EUA—value	4	9	7.20	7.00	1.22	0.168	0.006
EUA—truth social	5	9	7.80	8.00	1.24	0.264	<0.001
EUA—truth physical	3	9	7.08	7.00	1.46	0.187	0.001
SEUA—personal taste	9	15	10.70	10.50	1.02	0.254	<0.001
SEUA—aesthetic	10	15	10.68	10.00	1.35	0.442	<0.001
SEUA—value	7	15	11.88	12.00	2.04	0.134	0.067
SEUA—truth social	9	15	13.18	14.00	2.01	0.243	<0.001
SEUA—truth physical	5	15	12.33	13.00	2.57	0.154	0.018
nEUA—personal taste	6	9	6.55	6.00	0.71	0.329	<0.001
nEUA—aesthetic	6	9	6.55	6.00	1.04	0.452	<0.001
nEUA—value	3	9	7.30	7.50	1.36	0.196	<0.001
nEUA—truth social	4	9	7.95	9.00	1.34	0.309	<0.001
nEUA—truth physical	3	9	7.15	7.00	1.73	0.163	0.009
EUA—total score	28	42	34.90	35.00	3.04	0.116	0.186
nEUA—total score	27	43	35.50	36.00	3.78	0.153	0.020
SEUA—total score	46	71	58.75	59.00	5.75	0.142	0.040
Difference of profiles EUA–nEUA	0	4	1.15	1.00	1.12	0.228	<0.001
Difference of profiles nEUA–SEUA	0	2.5	0.66	0.50	0.52	0.247	<0.001
Difference of profiles EUA–SEUA	0	3	1.15	1.00	0.82	0.198	<0.001

extending the test was justified and brought noticeable improvement for the usability of the tool. The α values achieved in the SEUA version are still on the edge of acceptance for an instrument to be used in quantitative research; the commonly accepted lower level of acceptance is 0.7 (see George and Mallery 2003; Bland and Altman 1997). The values are too low to accept the tool as appropriate for assessing individuals and indicate that more work on this topic is required. It is worth noting that α values for judgements of personal taste and value were significantly lower in our research than the ones reported by Mason and Boscolo (2004), who obtained an α of 0.69 and 0.90, respectively.

In order to verify which test items may decrease the internal consistency in each judgement domain, values of Cronbach's α after the exclusion of items were calculated for SEUA.

In the personal taste domain after the removal of two pairs of sentences Cronbach's α will rise (both from EUA, the original version of the instrument). These pairs are: *Robin says the stew is spicy. Chris says the stew is not spicy at all* (α will rise from 0.439 to 0.570) and *Robin thinks weddings should be held in the afternoon. Chris thinks weddings should be held in the evening* (α will rise slightly from 0.439 to 0.453).

In the domain of aesthetics after the removal of two pairs of sentences Cronbach's α will rise. These pairs are: *Robin thinks the first painting they look at is better. Chris thinks the second painting they look at is better* (from EUA; α will rise from 0.852 to 0.861) and *Robin thinks that porcelain figures are the most beautiful. Chris thinks that glass figures are the most beautiful* (from SEUA; α will rise slightly from 0.852 to 0.854).

In the values domain after the removal of one pair of sentences Cronbach's α will rise: *Robin thinks the government should limit the number of children families are allowed to have to keep the population from getting too big. Chris thinks families should have as many children as they choose* (from EUA; α will rise from 0.662 to 0.674).

In truths about social world domain after the removal of one pair of sentences Cronbach's α will rise: *Robin thinks one book's explanation of why the Punic wars began is*

right. Chris thinks another book's explanation of why the Punic wars began is right (from EUA; α will rise from 0.775 to 0.805) (in the original tool this pair of statements concerns the Crimean war. We decided to use something more neutral in view of the current political situation, hence the Punic wars).

In truths about the physical world domain after the removal of one pair of sentences Cronbach's α will rise: *Robin agrees with one book's explanation of the origin of life on Earth. Chris agrees with another book's explanation of the origin of life on Earth* (from SEUA; α will rise from 0.745 to 0.759).

4.2.2 Stability

Summary scores and scores for each domain in EUA and nEUA were compared to assess the stability of levels of epistemological understanding; recall that EUA and nEUA consist of the very same items, only nEUA was administered as a part of SEUA. Wilcoxon's signed-rank test was used due to deviations from the normal distribution observed in variables. No significant differences between the scores in EUA and nEUA were observed (Table 8). Such an outcome is a sign of the stability of levels of epistemological understanding between measures and, at the same time, the stability of the original version of the tool designed to assess these characteristics.

In an attempt to assess if participants' profiles differed significantly between measures and test versions, the difference scores were calculated. For each judgement domain, one point was added to the difference score when the participant's score switched from one level to the previous or the next one (e.g. from A to M); between subsequent measurements and analogically for smaller (e.g. a half point for the switch from M- to A) and larger (e.g. two points for the switch from A to E) changes in the profile. Details of this scoring are given in Table 9.

The analysis revealed that for each version of the tool, some variability in profiles between measures was present. Average profile differences between EUA and SEUA and EUA and nEUA equaled 1.15, while the average difference

Table 7 The results of internal consistency analysis (using Cronbach's α)

	EUA	nEUA	SEUA
Personal taste	0.004	0.090	0.439
Aesthetics	0.776	0.861	0.852
Value	0.058	0.542	0.662
Truth: social	0.595	0.661	0.775
Truth: physical	0.323	0.648	0.745

Table 8 The results of Wilcoxon's signed-rank test

	EUA		nEUA		Z	p
	M	SD	M	SD		
Personal taste	6.40	0.67	6.55	0.71	-1.291	0.197
Aesthetics	6.43	0.87	6.55	1.04	-1.518	0.129
Moral values	7.20	1.22	7.30	1.36	-0.668	0.504
Truth: social world	7.80	1.24	7.95	1.34	-0.684	0.494
Truth: physical world	7.08	1.46	7.15	1.73	-0.277	0.781
Overall	34.90	3.04	35.50	3.78	-1.034	0.301

between nEUA and SEUA equaled 0,66. Bearing in mind that the difference of one point indicates a change from one level to an adjacent one (e.g. from A to M or from M to E), the average profile differences are not that large. However, the standard deviations compared to the average differences are quite high, indicating a noticeable variability in this measurement (see Table 10).

Subsequently, differences in profiles between EUA, nEUA and SEUA were compared. The Friedman test was used because of non-normality of distribution of analysed variables. The analysis confirmed that there are significant differences between the profile differences (Table 10). Post hoc analysis revealed that the difference of profiles between nEUA and SEUA was significantly smaller than the difference between EUA and SEUA ($p = 0.049$). This can be explained by the fact that the most similar profiles were constructed based on outcomes from the same (second) session. A lack of significant differences in the magnitude of profile changes between EUA and nEUA and between EUA and SEUA can serve as an indication of similar profile stability for nEUA and SEUA. Both yielded similar differences when compared to the EUA.

4.3 Additional Qualitative Data

Due to the fact that the study was carried out in the form of an interview, it was possible to gather important information concerning the instruments' content and its reception. Furthermore, experimenters were able to react immediately in cases of any misunderstanding and clarify the instructions.

During the interviews it turned out that some of the test items seemed to be more controversial than the others. One such item was: *Robin believes one mathematician's proof of the math formula is right. Chris believes another mathematician's proof of the math formula is right* (judgements of truth about the physical world; EUA). Responses and explanations given by participants suggest

that this item is so highly knowledge dependent, that it may not measure epistemological understanding in an appropriate way. The main issue was the fact that some of the participants were not familiar with the notion of the mathematical proof (e.g. suggesting that proof may be incorrect, which indicates that what they had in mind was probably a notion of a mathematical proof construed with a strong social flavour; see Ernest 1998, pp. 182–187). In some cases such as these, in order to make sure the response is a sign of an individual epistemic level that was employed and is not a result of a lack of knowledge, the experimenter briefly explained what a proof is and then noted if the subject changed his or her mind (only in the SEUA version of the instrument). In the final analysis, however, the first responses were used, as providing such explanations was employed only in the case of some of the participants. It was noted, that 6 participants, after the experimenters' explanations, changed their responses (five from A to M, and one from E to M). Another somewhat problematic item was: *Robin says the stew is spicy. Chris says the stew is not spicy at all* from EUA. Some of the participants pointed out in their explanation that the spiciness of the stew can be objectively measured (for example, on a Scoville heat scale). Other participants related the spiciness to the ingredients of the stew, arguing that if it contained spices like pepper or chili powder, the person that says that it is not spicy cannot be right. Both explanations bring attention to the fact that assessment of this pair of statements can be related to some objective measure, unlike in the case of other statements in the personal taste judgement domain.

Test items used in the instrument can be divided into two groups: pairs of abstract and concrete statements. Since the authors of the original tool did not mention these characteristics in their paper, it seems possible that they did not notice this. As concrete test items we consider sentences which include terms referring to specific objects (or features) of the external world as well as particular events or opinions on particular topics. Examples of such test items are: *Robin says warm summer days are nicest. Chris says cool autumn days are nicest* (EUA), *Robin thinks that porcelain figures are the most beautiful. Chris thinks that glass figures are the most beautiful* (SEUA), *Robin thinks lying is wrong. Chris thinks lying is permissible in certain situations* (EUA). Abstract test items, in contrast, include terms, which do not refer to specific objects, events or beliefs; their reference is rather a group or a class of some entities, not the entities as such (material or non-material). In the case of abstract test items, the only information the subjects got was that the references of crucial terms in pairs of statements are different (just different—not exclusive or complementary). Abstract test items are, for example: *Robin thinks the first piece of music*

Table 9 The calculation of difference scores

I	II	Scoring
A	A+, M–	0.5
	M	1
	M+, E–	1.5
	E	2
M	A+, M–, M+, E–	0.5
	A, E	1
E	M+, E–	0.5
	M	1
	A+, M–	1.5
	A	2

Table 10 The results of the Friedman test

	M	SD	F	df	p
Difference of profiles: EUA–nEUA	1.15	1.12	9.339	2	0.009
Difference of profiles: nEUA–SEUA	0.66	0.52			
Difference of profiles: EUA–SEUA	1.15	0.82			

they listen to is better. Chris thinks the second piece of music they listen to is better (EUA)—since the pieces of music they were thinking about were unknown; Robin thinks one book's explanation of why the Punic wars began is right. Chris thinks another book's explanation of why the Punic wars began is right (EUA)—since the exact explanations Robin and Chris referred to were not given; Robin has one view on the causes of unemployment. Chris has a different view on the causes of unemployment. (SEUA)—and, again, it was only known that their views are different. It should be noted, that in her earlier work Kuhn (1991) has discussed the tendency for individuals to be influenced by their own opinions on issues, even when prompted to discuss the possibility of making a judgement in a broader sense. Nevertheless, the authors of the original tool do not refer to these findings.

Some of the participants took notice of abstract-to-be test items saying that it was hard for them to choose an appropriate answer, whilst they did not know the exact objects Robin and Chris referred to. Sometimes the experimenter's suggestions (e.g. to think about such objects as different but not specific) were helpful. In a few cases, our subjects replied that without the information concerning the exact reference, they must indicate that Robin and Chris are equally right, but having more details they would be able to choose one option. Although it was emphasised that the task is to determine whether two views can be right at the same time, some subjects, having been exposed to concrete test items, had a tendency to pick the statement they found to be more appealing to them and tried to maintain this strategy in the case of abstract sentences.

The distribution of concrete and abstract statements was not even across the five judgement domains (see Table 11). This made the comparison of results obtained in concrete and abstract pairs impossible.

Out of the ten items added in SEUA, in comparison with EUA and nEUA, five were concrete and five were abstract.

Among all judgement domains, test items concerning values seemed to be the most problematic in terms of the subjects' inability to suppress personal preferences, which manifested themselves by choosing the more appealing view. As Table 11 indicates, the values domain is the only category which included only concrete pairs of sentences.

4.4 Analysis of Profile Patterns

The profile patterns observed in this study were mostly consistent with those reported by Kuhn et al. (2000). None of the participants were absolutists in the aesthetics and personal taste judgement domains, which is a finding appropriate for the studied group with respect to their age and education. In the domain of personal taste, only three participants (7.5 %) were classified as evaluativists for both the EUA and nEUA versions, and only one (2.5 %) for the SEUA version. Such a low percentage of switches from the multiplist to the evaluativist level supports the claim that in most people this transition never occurs.

Some controversies appear in the domain of values. In the original study, many participants failed to make a transition from the absolutist to the multiplist level. In our study, in the case of only one person, the transition from absolutist to multiplist did not happen (EUA: MMAMM, nEUA: MMAMM, SEUA: MMA+MM); and in one person the SEUA produced a pattern in which in the value domain result was A+, while in both EUA and nEUA the result was M (EUA: MMMEA, nEUA: MMMEA, SEUA: MMA+EM).

We also found that several subjects exhibited patterns inconsistent with Kuhn et al.' hypothesis of the order of transitions, such as: evaluativist in the judgement domain

Table 11 Distribution of concrete and abstract test items

Judgement domain	Concrete	Abstract
EUA and nEUA		
Personal taste	2	1
Aesthetics		3
Moral values	3	
Truth about the social world		3
Truth about the physical world		3
EUA version—summary	5	10
SEUA		
Personal taste	4	1
Aesthetics	1	4
Moral values	5	
Truth about the social world		5
Truth about the physical world		5
SEUA version—summary	10	15

of truths in the social world and absolutist, or multiplist in the domain of truths in the physical world (27.5 % in both EUA and nEUA, 20 % in SEUA). A significantly lower number of participants exhibited patterns inconsistent with the hypothesised transformation trend in the personal taste and aesthetics domains (5 % in EUA and 2.5 % in nEUA).

5 Discussion

Introducing the suggested modifications to the instrument resulted in a remarkable increase in the reliability measures. In almost every domain, SEUA exhibits higher reliability than EUA or nEUA. This makes SEUA more suitable for quantitative research than the original version of the instrument. Further improvement in reliability measures is possible, as indicated in the analysis of the internal consistency of SEUA. Furthermore, SEUA is as stable as the EUA, as the participants profiles do not change during examination with SEUA compared to the EUA. However, SEUA offers more variability in terms of content, refers to more aspects in each domain which can provide better ecological accuracy and decrease the potential dependence of answers on the specific content of the test item and not the judgement domain itself. As a result, SEUA can serve as an improved alternative instrument for assessing an epistemological level of understanding of an individual.

The quantitative scoring method allows for the assessment of the internal consistency of each subscale of the tool and checking score stability over time. Moreover, it enables more comprehensive comparisons of the results of different people. Another important modification is the change in the form of research from paper–pencil to interview. This kind of interaction brings about a new set of data. An experimenter can observe the reactions of a participant, if he or she can correctly understand the instruction or has any doubts. If so, the immediate answer or clarification can be provided.

While many of those modifications were present in previous research, our proposal integrates the most important changes that improve the instrument. The modifications addressed most controversies that could have influenced the reliability and applications of the tool.

The results obtained in the study in regard to profile patterns are mostly consistent with the hypothesised order of transitions between levels of epistemological understanding. We did not observe the same difficulties with the values domain as Kuhn et al. (2000), which can be related to the fact that, unlike in the original study, our subjects formed a rather homogenous group. The fact that some of the participants from the study of Kuhn et al. tended to hold on to absolutist views in the moral category, can also be a

result of the specific characteristic of this domain—some research suggest that judgements concerning morality “cannot be reduced to matters of personal preferences or factual beliefs” (Krettenauer 2004, p. 462) and therefore the values domain is not a truly epistemic one. This issue can give rise to the question if the values domain should be excluded from the tool. In the current version, we opt for keeping it, as SEUA is just an adaptation of the tool by Kuhn et al. (2000). SEUA was constructed based on a certain theoretical model of personal epistemology, and it requires further research to determine, if excluding the values domain is justified. We suggest that the problem of concrete versus abstract statements should be addressed in the first place (see Sect. 5.1).

The homogeneity of the tested group probably resulted in low variation in levels showed by participants in the personal taste and aesthetics domains. Deviations from the hypothesised transition order were observed in the shifts in both truth domains, suggesting that for some people, these transformations are more complicated. Some of the issues with profile patterns can be connected to the disproportions of concrete and abstract test items in certain domains and a lack of abstract items in the values domain.

5.1 Limitations of the Study and of the SEUA Test Tool

SEUA, the extended tool for assessing levels of epistemological understanding, despite being more stable and reliable than the original EUA version, has some limitations that can influence its use. The main disadvantage is that it is not suited for testing more than one person at the time. Performing the test requires approximately 30–40 min per person. While the long testing time contributes to difficulties in studying larger samples, the direct interaction with the researcher can be a source of discomfort from the participant’s point of view. Testing in this setting can be more stressful than a paper-and-pencil test, as the answers have to be given out loud along with explanations, with no preparation. The recording of the sessions can be another factor that negatively influences the participant’s mood. In the future, the possibility of creating a paper–pencil version of SEUA should be explored, along with precise instructions for the experimenters. While this can be difficult due to the inevitable loss of information that is provided by the participant–experimenter interaction, a paper version will be crucial for conducting research on bigger samples more effectively.

The distinction between concrete and abstract test items gives rise to the questions of whether such heterogeneity affects answers and whether it is possible to recreate the instrument to contain only concrete or only abstract statements. During the interviews it was observed that some

subjects, despite being given instructions only to evaluate if two statements could be right and not to share their own views on the issues raised, in evaluating concrete test items were pointing at answers which they considered to be true. A separation of one's own view from the general question of rightness was impossible. Addressing this issue might be a topic for future analysis. As one of the possible ways to inspect the effect of abstractness of test items on the levels of epistemological understanding we propose a comparison of the results from alternative versions of the instrument with balanced concrete to abstract items ratio. Another, more radical option is the removal of all of the concrete items. This approach must be followed by constructing more abstract pairs of statements in order to maintain the satisfactory length of the tool. We suggest that maybe the controversy of concrete and abstract items can be solved by transforming the concrete statements to abstract, without changes in their general topic. For example, a pair of concrete judgements about personal taste:

Chris says cool autumn days are nicest.

Robin says warm summer days are nicest.

can be transformed into

Chris says one season has the nicest weather.

Robin says another season has the nicest weather.

The version of the tool that will be an effect of such changes should be tested for its reliability and stability to see if the tool can be improved that way.

It has been argued that the assessment methods based on the model described by Kuhn et al. (2000) may understate the actual number of absolutists (e.g. Ahola 2009; Barzilai and Weinstock 2015). It cannot be ruled out, that some people may find it harder to reply in an absolutist fashion when they are exposed to abstract test items. Nevertheless, introducing the interview method by providing justifications to each given answer and allowing the researcher to clarify any possible ambiguities, could increase the chance that the obtained scores will adequately reflect the level of epistemological understanding of a subject. Furthermore, in our study the participants formed a homogenous group of people from whom—according to the model proposed by Kuhn et al. (2000)—one would expect a small percentage of absolutists.

5.2 Future Research

Further research is needed in order to provide an instrument that not only has satisfactory psychometric properties, but is also easy to administer. The ideal goal should be a shortened, paper-and-pencil or computer test that can be carried out simultaneously on many subjects. A computer administered test might offer some advantages, because of

the possible control of a participant's returning to test items and changing the answers, as well as because of an automatic score calculation. Before creating paper-and-pencil or computer version of the tool, the issues highlighted in the Sect. 4 should be addressed. Ideally, a new version of the tool, with proper adjustments, should be constructed for interview form and analysed prior to transforming it to paper-and-pencil version.

Another important issue is improving the scoring technique in order to provide quantitative outcomes useful for group studies as well as, possibly, more comprehensive information about an individual's level profile of epistemological understanding. As for plans to create versions of the tool with equal numbers of concrete and abstract items, the question arises if one can propose abstract versions for the values judgement domain. No such question was introduced in the original instrument and none was added during our modifications, suggesting that this is a problem which can be the source of differences between values and other judgement domains. Additionally, the relationship between the proportion (or lack) of concrete items to abstract ones and answers in the value judgement domain should be carefully examined, so the tool could be further improved by removing this domain or modifying it.

Acknowledgments Research reported in this paper were supported by the National Science Centre, Poland (DEC-2013/10/E/HS1/00172).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ahola S (2009) Measurement issues in studying personal epistemology. *Psychol Soc* 2(2):184–191
- Barzilai S, Weinstock M (2015) Measuring epistemic thinking within and across topics: a scenario-based approach. *Contemp Educ Psychol* 42:141–158
- Barzilai S, Zohar A (2009) The role of epistemic thinking in online learning. In: Eshet Y, Caspi A, Eden S, Geri N, Yair Y (eds) *Learning in the technological era: proceedings of the 4th Chais conference on instructional technologies research*. The Open University of Israel, Ra'anana, pp 29–33
- Baxter Magolda MB (2004) Evolution of a constructivist conceptualization of epistemological reflection. *Educ Psychol* 39(1): 31–42
- Bland JM, Altman DG (1997) *Statistic notes: Cronbach's alpha*. *Brit Med J* 314:572
- Christodoulou A, Osborne J, Richardson K, Howell-Richardson C, Simon S (2010) A study of student beliefs about the epistemology of science and their relationship with students' personal epistemologies. Paper presented at the NARST conference, Philadelphia, 21–24 Mar 2010

- Ernest P (1998) *Social constructivism as a philosophy of mathematics*. State University of New York Press, Albany
- George D, Mallery P (2003) *SPSS for Windows step by step: a simple guide and reference*, 4th edn. Allyn & Bacon, Boston
- Hofer B, Pintrich P (1997) The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Rev Educ Res* 67:88–140
- Kitchener KS, King PM (1981) Reflective judgment: concepts of justification and their relationship to age and education. *J Appl Dev Psychol* 2(2):89–116
- Krettenauer T (2004) Metaethical cognition and epistemic reasoning development in adolescence. *Int J Behav Dev* 28:461–470
- Kuhn D (1991) *The skills of argument*. Cambridge University Press, Cambridge
- Kuhn D (1999) A developmental model of critical thinking. *Educ Res* 28(2):16–46
- Kuhn D, Cheney R, Weinstock M (2000) The development of epistemological understanding. *Cogn Dev* 15:309–328
- Mason L, Boscolo P (2004) Role of epistemological understanding and interest on interpreting a controversy and in a topic-specific belief change. *Contemp Educ Psychol* 29(2):103–128
- Mason L, Scirica F (2006) Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learn Instr* 16:492–509
- Mason L, Boldrin A, Zurlo G (2006) Epistemological understanding in different judgement domains: relationships with gender, grade level, and curriculum. *Int J Educ Res* 45(1–2):43–56
- Perry WG Jr (1970) *Forms of intellectual and ethical development in the college years: a scheme*. Holt, Rinehart and Winston, New York
- Piaget J (1950) *Introduction a l'epistemologie genetique*. Presses Universite de France, Paris
- Tabak I, Weinstock M (2008) A sociocultural exploration of epistemological beliefs. In: Khine MS (ed) *Knowing, knowledge and beliefs epistemological studies across diverse cultures*. Springer, Dordrecht, pp 177–195