The Development of Epistemological Understanding Revisited: Enhancing Reliability of the Tool by Using Only Abstract Items

Natalia Żyluk¹, Karolina Karpe², and Mariusz Urbański¹*  

Abstract
The aim of this paper is to describe the process of modification of the research tool designed for measuring the development of personal epistemology—Standardized Epistemological Understanding Assessment (SEUA). SEUA was constructed as an improved version of the instrument initially proposed by Kuhn et al. SEUA was proved to be a more reliable instrument than its predecessor; however, further changes were necessary to obtain better reliability and easier to administer form. During further research, we observed that test items used in this tool could be divided into abstract and concrete, which were approached differently by a subset of our participants. In their cases, the inability to suppress personal preferences in responding to concrete items threatened the tool’s validity, as the instrument may measure preferences rather than epistemological beliefs in this situation. SEUA was therefore modified to create a full-abstract version (SEUA-A). Both versions were administered in an online form. The performance of two versions of the tool was compared. The study results allow us to conclude that our online SEUA-A, which consists of only abstract items, is the most reliable version of the tool.

Keywords
personal epistemology, epistemological understanding, tool development, SEUA-A

Introduction
The main aim of this paper is to describe the process of modification of the research tool designed for measuring the development of personal epistemology—Standardized Epistemological Understanding Assessment (SEUA). SEUA was constructed as an improved version of the instrument initially proposed by Kuhn et al. (2000). The instrument by Kuhn et al. (2000) was designed to account for the domain-dependent coordination of objective and subjective dimensions of knowing. This tool has been used in research on, for example, the relationship between epistemological beliefs and engagement in argumentative discourse as well as shifting an attitude within a digital dialogue game (Noroozi, 2016); the role of epistemic thinking in online learning processes (Barzilai & Zohar, 2009); relations between epistemic beliefs and achievement goal profiles (Madjar et al., 2017); the influence of epistemological understanding and interest in the interpretation of controversial text and topic-specific belief changes (Mason & Boscolo, 2004); or sociocultural determinants of epistemological understanding (Tabak & Weinstock, 2008).

In developing SEUA, we suggested (Żyluk et al., 2016) several changes to the original instrument by Kuhn et al. (2000). The proposed modifications addressed most controversies that could have influenced the tool’s psychometric properties. The most significant changes included extending the list of test items (from 15 to 25), a new administration procedure (from paper-pencil method to a structured interview) and the introduction of a quantitative scoring method. Some of these modifications had already been introduced in previous studies, for example, Christodoulou et al. (2010) as well as Mason and Boscolo (2004), who introduced
alternative scoring methods; Mason et al. (2006) in their research applied elements of interviewing, that is, asking a limited number of participations for oral clarification of their answers. In our study, however, we have decided to integrate them for the sake of developing the tool that would comprehensively meet our predefined objectives (see the detailed discussion in Żyluk et al., 2016).

The outcome—SEUA—was proven to be more reliable than the original instrument by Kuhn et al. (2000), but—as indicated during the analysis of the internal consistency—further improvement in reliability measures remained possible.

The fact that the study employing SEUA was carried out interactively, which encouraged our participants to share their thoughts enabled us to gather important information concerning the tool’s content and its reception. Furthermore, it made it possible to react immediately in case of any misunderstanding(s) and provide clarification whenever needed. One of the observations we made in the course of conducting interviews was that test items used in SEUA can be divided into two groups: abstract (those that refer to a general class of objects, features, or beliefs) and concrete (those that refer to particular objects, object properties, or beliefs). This distinction remained unnoticed until the study began as participants’ comments and reactions were factors that drew our attention to its presence. In our previous article (Żyluk et al., 2016) we discuss these remarks by pointing out that for some subjects, it was hard to follow the tool’s instructions while evaluating concrete test items. The main issue, in this case, was participants’ inability to suppress personal preferences, which can modify the answering process and, consequently, substantially affect the validity of the tool (in such a way that the instrument may not measure epistemological beliefs any longer, at least not exclusively).

The second source of our doubt concerning the presence of concrete test items in the SEUA instrument were the results of the pilot study of a paper-pencil version of SEUA (Żyluk et al., 2017) in which the instrument exhibited lower reliability measures than in a structured interview setting. We suspect that this result was caused by a lack of support from the researcher during the responding process, which may result in a misunderstanding of instructions. It is fair to say that in the interview study, the support of the researcher was particularly needed in the context of concrete test items.

Drawing on the observations gathered, we decided to modify the tool even further and create a full-abstract version of SEUA (referred to as SEUA-A) and to compare it with regular SEUA, both in an online form. The outcome of these processes is the focal point of this paper.

This paper is structured as follows: the Background section introduces the basic theoretical assumptions that underlie described tools for measurement of personal epistemology (individuals’ theories on knowledge and knowing) and provides an outline of Kuhn et al. (2000) original tool, as well as its improved form—SEUA (Żyluk et al., 2016); limitations of these tools, such as abstract-concrete distinction and a rationale behind developing a new version are also described. We then characterize the changes introduced to SEUA that resulted in obtaining SEUA-A and describe the research process that would allow us to compare the regular and abstract versions of SEUA (Method section). The Results section provides information regarding descriptive statistics, reliability measures, and comparisons between the performance of tool versions. We close with the discussion and recommendations for further use of SEUA and SEUA-A.

Background

The study of individuals’ theories on knowledge and knowing (personal epistemology) has become an important line of inquiry within fields of developmental psychology and educational studies. The origins of empirical research on personal epistemology are associated with Perry’s (1970) longitudinal interview studies, yet its very roots can be found in Piagetian genetic epistemology and stage theory of development (1950). As pointed out by many scholars (e.g. Ahola, 2009; Braten, 2010; Hofer & Pintrich, 1997; Labbas, 2013; Sandoval et al., 2016), a variety of different approaches to studying personal epistemology exists (with various synonyms used, e.g., epistemological understanding, epistemological assumptions, and epistemic thinking; Holma & Hyytinen, 2015). Consequently, many measures are built upon the different conceptualization of the central construct. The instrument by Kuhn et al. (2000) was based on a model of personal epistemology, according to which the epistemological progression of an individual is dependent upon the coordination of the subjective and objective aspects of knowledge and knowing. In this view, personal epistemology is considered a system of beliefs that is developmental and processual (vs. static) in its nature (Hofer, 2001). The development of mature epistemological understanding starts with the domination of objectivism, which is subsequently replaced with subjectivism. Finally, both dimensions of knowing may be integrated (Kuhn et al., 2000). In reference to these phases, Kuhn et al. (2000) distinguished between four levels (stages, stances) of epistemological understanding: realist and absolutist (objectivism), multiplist (subjectivism), and evaluativist (integration of objectivism and subjectivism).

Individuals at the first two stages (realist and absolutist level) see reality as being directly accessible and “knowable.” They consider knowledge as certain and coming from a source external to the subject. However, what is essential is that realists and absolutists have different views on the nature of assertions, which can be treated either as copies of reality at a realist level or as true or false facts that represent reality correctly or incorrectly, at an absolutist level. Given the latter, under absolutist interpretation, it is possible for a belief to be false. Under typical development, a realist stance is
present only with small children, so neither the original instrument nor SEUA assessed epistemological understanding at a realist level.

For an absolutist, when people have different views on a particular topic, it is not possible that both of them are right (as there can be only one “objective” reality they can refer to). To transfer from an absolutist to a multiplist level, it is necessary to realize the subjective aspect of knowledge and its uncertainty.

People at the multiplist level begin to see reality as not directly “knowable” and treat knowledge as something that does not come from objective, external reality but as generated by human minds. Under a multiplist view, knowledge is tied with one’s point of view. Given the above, multiplists consider knowledge uncertain and lacking an objective dimension. For multiplists, there is the possibility that multiple accounts on a certain topic can be equally right: they treat assertions merely as opinions.

The most advanced and last stage of epistemological development is the evaluativist level. At this stance, both subjective and objective dimensions of knowledge are integrated. From evaluativists’ point of view, like multiplists, the reality is not directly “knowable,” and knowledge is uncertain. This reflects the objective side of this level, as it is possible for one view to have more merit than the other given the empirical evidence or soundness of the argumentation. For evaluativists, different accounts on a particular subject matter can have righteness simultaneously. Still, some can have more merit or be better justified than others (they treat assertions as judgments that can be an object of evaluation).

For the purpose of this article, only a general overview of levels of epistemological understanding is provided. For a more detailed description, the reader should consult Kuhn’s original papers on the topic of personal epistemology (e.g., Kuhn et al., 2000).

The problem of domain-generality versus domain-specificity is one of the most vital issues within the personal epistemology research field (e.g., see overview in Muis et al., 2006). Kuhn et al. (2000) advocated the view according to which levels of epistemological understanding need to be considered within certain judgment domains: personal taste judgments, esthetic judgments, value judgments, and truth judgments (with this last one further differentiated into two categories: judgments of truth about the social world and judgments of truth about the physical world). It was hypothesized that the first transition (from absolutist to multiplist) occurs first within the domain of personal taste judgments, then in esthetic judgments, later in the domain of value judgments, followed by the judgments of truth about the social world, and judgments of truth about the physical world. The second transition (from multiplist to evaluativist) takes place in reverse order. To verify if the epistemological progression occurs in a hypothesized order, Kuhn et al. (2000) developed a paper-pencil test tool which we decided to modify further (obtaining SEUA).

Standardized Epistemological Understanding Assessment (SEUA)

In this section, we will describe SEUA with a particular emphasis put on the main changes we introduced to the original instrument by Kuhn et al. (2000) while creating our tool (subsection 2.1.1). For clarity, we decided not to describe the instrument developed by Kuhn et al. (2000) separately, that is, that the general description of its modifications should suffice to provide its general outlook.

This part will also cover the central issue of the paper—a distinction between abstract versus concrete items (subsection 2.1.2), which, after it had emerged during the interview research with SEUA, inspired us to modify this tool further resulting in SEUA-A.

Tool description. The general idea behind the tool by Kuhn et al. (2000), which we preserved in SEUA, is to measure the levels of epistemological understanding by assessing the views on the relation of simultaneous rightness of two incompatible judgments. SEUA consists of 25 pairs of sentences—five for each judgment domain (personal taste judgments, esthetic judgments, value judgments, judgments of truth about the social world, and judgments of truth about the physical world). Each pair consists of two conflicting positions on an issue, presented by two people: Chris and Robin. In the case of every pair, a subject needs to answer the questions posed to determine if a transition from one level of epistemological understanding to another has taken place. To assess if the transition from the absolutist to the multiplist level has occurred, participants are asked whether only one presented position could be right, or both could have some “rightness.” The diagnostic answer for the absolutist level is “Only one view can be right.” If participants chose the option “Both could have some “rightness’,” they are then asked whether one view could be better or more right than the other. The answer “One could not be more right than the other” serves as an indicator for the multiplist level, while “One could be more right” is the diagnostic answer for the evaluativist level. To summarize the questioning process: when the participant exhibits absolutist beliefs for a given pair of sentences, they answer only one question for this item, that is, to specify if somebody is multiplist or evaluativist, one additional question needs to be asked.

Exemplary test items from each of the judgment domains were as follows:

Judgments of personal taste:
Chris says cool autumn days are the nicest.
Robin says warm summer days are the 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.

In the tool developed by Kuhn et al. (2000), the number of test items was 3 for each judgment domain (15 overall). We extended the length of the tool by adding two new pairs of sentences in each domain. Expanding the list of items was mainly directed at lowering the probability of the overall score being heavily influenced by the particular thematic content of items.

In SEUA, as well as in the original test instrument, a subject was assigned a category (an absolutist—A, a multiplist—M, or an evaluativist—E) in each of the judgment domains, depending on the number of answers that fall into a specific category. In the original tool the participant was categorized as absolutist, multiplist, or evaluativist in the particular judgment domain if, for at least two out of three statements in a particular domain, they reply in a way characteristic for one of these levels of epistemological understanding. Having three different answers for specific judgment domain resulted in receiving the category of multiplist for this domain.

In the case of SEUA, the scoring method was slightly modified in such a way as to take into account an extension of the number of test items. We decided not to transfer the scoring strategy used in the 15-item-long tool, as we consider the proportion of 3 to 5 items as too weak to justify the assignment of a specific category (for some exceptions, see below). In addition to the original three letters, the signs + and − were used in describing the obtained levels. As a result, the complete list of possible categories was the following: A, A+, M−, M, M+, E−, and E. This solution provides a more fine-grained assessment of epistemological views held by an individual, as it allows one to capture some “intermediate” stances. A somewhat similar scoring strategy was introduced in work by Christodoulou et al. (2010), in which the authors introduced additional categories (A+, E− as well as “Indeterminate”) to represent the variability of subjects’ responses. Despite seeming similarities, our scoring method was developed as an improvement of Kuhn et al. (2000) independently of the method proposed by Christodoulou et al. (2010) and is not an adaptation or modification of their scoring.

This scoring method is summarized in Table 1 in columns i to iv (columns (i)–(iii) refer to the number of answers characteristic for particular level; column (iv) presents the category ascribed on the basis of provided answers; the content of column (v) will be addressed later on). Providing at least four answers that fell into a specific category resulted in receiving a “clear” A, M, or E category with no additional signs. There were two exceptions to this rule, as we also decided that a “clear” multiplist category will be assigned in the cases when somebody received the 2−1−2 and 1−3−1 pattern for A, M, and E, respectively (as we thought these distributions reflect the intermediacy of a “clear” M).

As the tool proposed by Kuhn et al. (2000) was designed to measure domain-specific epistemic beliefs, there was no possibility to obtain a single summary score reflecting the general level of epistemological understanding. Instead, a participant could be assigned a specific profile, for example, AMMEE, where consecutive letters represent levels of epistemological understanding in the domains of personal taste, aesthetics, value judgments, truth judgments about the social world and truth judgments about the physical world, respectively.

In the case of SEUA, we preserved this strategy, but, as previously mentioned, apart from that, we introduced a quantitative scoring method which allowed us to assess the internal consistency of each subscale (understood as a set of items concerning a single judgment domain) of the instrument as well as to detect some of the tool’s weaker points (e.g., test items that negatively correlate with the rest of the items within one domain). Aside from a qualitative profile, participants were assigned points for their answers. The points were summed up within each of the five domains. For every answer A, 1 point was provided; for M, 2 points; and E, 3 points. For SEUA, the scores within domains ranged from 5 to 15. The last column (v) of Table 1 presents quantitative scores that can be obtained depending on the responses given within one judgment domain (five items). As most of these scores repeat for two or more different qualitative letter category-labels, they should be reported as a pair of category-score for the results to be informative.

In the study described in Żyluk et al. (2016), aside from summing up the points within each domain separately, for descriptive purposes only, we also decided to calculate summary scores for the instrument as a whole (for a maximum of 75 points). However, given the domain dependency of epistemological beliefs, we found this general score not as directly interpretable; in particular, it is not delivering any reliable information about the general level of the epistemological understanding of an individual.

The research with SEUA had the form of a recorded interview. During the study, participants were presented with pairs of statements in a written form, each pair on a separate card (the experimenter read the pair first aloud and then gave the card to the participant). Questions concerning the rightness of the presented judgments (see above) were asked by the experimenter and repeated for every test item. To ease the answering process, each participant received a response schema: a paper sheet presenting which questions should be answered and in what order for each pair of sentences (all the materials used in the research with SEUA in an interview
Table 1. Assignments of the Categories and Corresponding Scores in SEUA.

| (i) | (ii) | (iii) | (iv) | (v) |
|-----|------|-------|------|-----|
| A   | M    | E     | Category | Score |
| 5   | A    | 5     | A      | 5    |
| 4   | A    | 6     | 6      | 6    |
| 4   | A    | 7     | 7      | 7    |
| 3   | A+   | 8     | 8      | 8    |
| 2   | M−   | 9     | 9      | 9    |
| 1   | M    | 10    | 10     | 10   |
| 2   | M    | 11    | 11     | 11   |
| 1   | M    | 12    | 12     | 12   |
| 4   | E−   | 13    | 13     | 13   |
| 1   | E    | 14    | 14     | 14   |
| 5   | E    | 15    | 15     | 15   |

Abstract-concrete distinction. The main issue that attracted attention as a possible area of improvement was the division between abstract and concrete items, which seemed to interfere substantially with answering process of some participants. An item is considered “abstract” when neither of the two statements in the pair refers to a concrete object, object property, event, or belief but instead to the general class of objects, properties, or beliefs. Correspondingly, a “concrete” item is an item where the statements refer to a known, specific object, property, or belief. To give some examples, Robin says warm summer days are the nicest. Chris says cool autumn days are the nicest (test item from Kuhn et al., 2000), and Robin thinks that porcelain figures are the most beautiful. Chris thinks that glass figures are the most beautiful. (new item introduced in SEUA) are concrete items, while Robin thinks the first piece of music they listen to is better. Chris thinks the second piece of music they listen to is better. (test item from the original instrument) and Robin has one view on the causes of unemployment. Chris has a different view on the causes of unemployment. (new item introduced in SEUA) are abstract items. As can be seen, whilst we know which season of the year or type of material the figures are made of were more preferred by Chris or Robin, the pieces of music they were thinking about remained unknown, and the explicit opinions on unemployment they referred to are not given. In abstract test items, in general, the only information the subject has got is that the references of crucial terms presented in these statements (e.g., pieces of music, views on the causes of unemployment) are different (just different—not necessarily exclusive or complementary).

As an abstract-concrete distinction was not mentioned in Kuhn et al. (2000), it seems that it first became apparent in the interview approach taken in Żyluk et al. (2016). It was noted before, however, that individual’s preferences on a given topic can influence their ability to make broader judgments (Kuhn, 1991), and what was observed in interviews performed during our previous study could provide some indication as to whether that was indeed the case.

It should be remembered that the general idea behind the tool we used in our research was to measure levels of epistemological understanding by assessing the views on the relation of simultaneous “rightness” that holds (or not) between two incompatible judgments on the same topic. As mentioned in the introduction, we observed that the presence of concrete test items in the tool might affect understanding of instructions, making some subjects more likely to understand their task as not to answer the question “Can only one of their views be right, or could both have some rightness?” but rather “Which one of their views is right, or do both have some rightness?” The difference between these two instructions may be seen as subtle, but its proper understanding seemed to be crucial to guarantee the tool’s validity. The actual task required considering the presented pairs of sentences in a more metaepistemic manner to assess the relation between them, regardless of the view of a subject on this particular subject matter. It is not to say that the content of these statements does not play a role at all, but it was important to the degree it determined the general topic Chris and Robin discussed.

In our research with SEUA (Żyluk et al., 2016), some participants, when faced with concrete items required interviewer’s intervention to follow the instructions: to judge the possibility whether both judgments can have some rightness at the same time, rather than decide if they agree with one of the statements. In a few cases, it was explicit that when the participants were exposed to concrete examples of Chris’ and
Robin’s beliefs, they could not distance themselves from their own views on the issues introduced in test items, even with the interviewer’s guidance. It was noted that statements from the values domain seemed to be the most problematic in terms of the subjects’ inability to suppress their personal preferences, which manifested itself in choosing the more appealing views. This problem was generally typical for concrete items (since they contained the views from which the subject could choose). Still, occasionally this tendency manifested itself to some extent for abstract items, too, as some subjects tended to react to abstract test items saying that it was hard for them to pick an appropriate answer since they did not know the items’ exact reference. Sometimes the researcher’s suggestions (e.g., to think about such objects as different but not specific) were helpful in these situations. In some cases, however, participants replied that with the lack of information of the exact item’s reference they must indicate that both views are equally right, but after its further specification, they would choose one option. We suspected that problems with the need for specification of the items’ reference in the case of abstract items were caused by the presence of concrete items in the tool, but we did not directly test this assumption empirically. We consider the problems with understanding the general question of the rightness of two different views as distinctive for concrete items. However, it remains possible that we did not note every single instance of instructions misunderstanding related to abstract-concrete distinction since not every participant tended to propose the justification of provided answer in a way that allowed to determine if they encountered this problem or not.

Given all of the above, it can be assumed that for a subset of individuals, the SEUA tool does not only measure their levels of epistemological understanding but the overall score for every domain is also influenced by individuals’ opinions on the test content.

Concrete items constitute 40% of the SEUA tool: 10 out of 25 items. In the original tool by Kuhn et al. (2000), the distribution between concrete and abstract test items was 5 out of 15. Of the 10 items we added in SEUA, 5 were abstract and 5 were concrete (this proportion was unintentionally obtained). As the number of concrete and abstract items is not balanced throughout the five domains, each domain score depends on concrete items to a different degree (see Table 2 for detailed data). Therefore, it is sound to develop the tool further to obtain a more robust measure of the levels of epistemological understanding that is not influenced by the presence of proportion of concrete items.

In our previous research (Żyluk et al., 2016), SEUA exhibited mostly satisfactory levels of internal consistency (up to Cronbach’s alpha coefficient of .852) with the significant exception of the domain of personal taste, where alpha reached the level of .439 (for the results for all of the subscales, see Table 6 or consult the paper cited above). We noted that while the values we obtained for SEUA are still higher than for the original version of the tool, more work oriented toward increasing reliability levels is required. In the research described in the paper by Żyluk et al. (2017), we used the paper-pencil version of SEUA and detailed instruction on how to understand the task correctly. The internal consistency measures obtained in this research exhibited a slightly lower Cronbach’s alpha coefficient level than in an interview version of the tool (see Table 6). We suspect that abstract-concrete heterogeneity may significantly affect the tool’s reliability in both cases. Consistent with this observation is that in our second research, we did not assist the participant while solving the task; therefore, the instructions might not have been understood correctly despite their thoroughness.

In the paper by Żyluk et al. (2016), we proposed a few possible ways to address the issue of the abstractness of test items and the effects of its presence in the tool on the assessment of levels of epistemological understanding. Two of our suggestions assumed either the removal of all of the concrete items (and creating more abstract pairs of statements to sustain the appropriate length of the tool) or transforming the existing concrete statements into the abstract. In these two cases, the results obtained with the new instrument should be compared with those collected using the regular version of SEUA in terms of their reliability measures. Eventually, we decided to lean on the second approach.

### Method

#### Tool’s Modification Procedure

As we strived to obtain a homogenous tool regarding the item form, the main body of work in this round of improvements to the SEUA tool consisted of rephrasing the concrete items into abstract ones. We started with identifying all of the concrete items included in SEUA. As mentioned earlier, an item was considered “concrete” when the statements referred to a known, specific entity (object, property, or belief). In the next stage, to assure the correctness of the selection, the process of identifying concrete items was performed again by competent judges—a group of 10 students who finished university-level psychometric training. Subsequently, the same group of students served as competent judges for the

### Table 2. Distribution of Concrete and Abstract Test Items in SEUA.

| Judgment domain               | Concrete | Abstract |
|-------------------------------|----------|----------|
| Personal taste                | 4        | 1        |
| Esthetics                     | 1        | 4        |
| Moral values                  | 5        | 0        |
| The truth about the social world | 0        | 5        |
| The truth about the physical world | 0        | 5        |
| SEUA—overall                  | 10       | 15       |

One interesting problem that manifested itself to some extent for abstract items, too, is some cases, however, participants replied that with the lack of information of the exact item’s reference they must indicate that both views are equally right (since they contained the views from which the subject could choose). Still, occasionally this tendency manifested itself to some extent for abstract items, too, as some subjects tended to react to abstract test items saying that it was hard for them to pick an appropriate answer since they did not know the items’ exact reference. Sometimes the researcher’s suggestions (e.g., to think about such objects as different but not specific) were helpful in these situations. In some cases, however, participants replied that with the lack of information of the exact item’s reference they must indicate that both views are equally right, but after its further specification, they would choose one option. We suspected that problems with the need for specification of the items’ reference in the case of abstract items were caused by the presence of concrete items in the tool, but we did not directly test this assumption empirically. We consider the problems with understanding the general question of the rightness of two different views as distinctive for concrete items. However, it remains possible that we did not note every single instance of instructions misunderstanding related to abstract-concrete distinction since not every participant tended to propose the justification of provided answer in a way that allowed to determine if they encountered this problem or not.

Given all of the above, it can be assumed that for a subset of individuals, the SEUA tool does not only measure their levels of epistemological understanding but the overall score for every domain is also influenced by individuals’ opinions on the test content.

Concrete items constitute 40% of the SEUA tool: 10 out of 25 items. In the original tool by Kuhn et al. (2000), the distribution between concrete and abstract test items was 5 out of 15. Of the 10 items we added in SEUA, 5 were abstract and 5 were concrete (this proportion was unintentionally obtained). As the number of concrete and abstract items is not balanced throughout the five domains, each domain score
procedure of generation and the selection of alternative, abstract variations of items. The abstract items were generated from concrete ones by referring to the same general category and indicating that Chris and Robin hold different views, without informing what precisely those views are. For example, in the item *Robin says warm summer days are the nicest and Chris says cool autumn days are the nicest* we can see that it explicitly refers to known, specific notions of summer and autumn days, making this item concrete and introducing the risk that participants will choose which days they prefer, instead of thinking about the possibility of the simultaneous “rightness” of Chris’ and Robin’s views. In the abstract version, we referred to the more general notion of the season and only informed that Chris’s and Robin’s views differ from each other, not including the information on how they differ. Therefore, the final, abstract version is as follows: *Robin says that a certain season is the nicest and Chris says that another season is the nicest*. An analogous procedure was applied to the rest of the concrete items, with one exception.

In one case, we opted for the complete change of the item’s content since satisfactory transformation from the concrete to the abstract version was not possible. The item in question is *Robin thinks people should take responsibility for themselves and Chris thinks people should work together to take care of each other*. The item that replaced this one in moral values is: *Robin has certain opinion on the use of animal products and Chris has a different opinion on the use of animal products*. We found it a fair exchange, as the use of animal products is a recognized moral dilemma of the contemporary world and an issue which participants are likely to be familiar with.

The minor changes involved rephrasing the test items that were ambiguous or not clear for a significant number of participants. While conducting the interviews, it was noticeable that some participants struggle with responding to several items more than to others. For two items, the reason for such hardship seemed to be precisely the content of the items. First of the troublesome items was *Robin says the stew is spicy and Chris says the stew is not spicy at all* from personal taste domain. As discussed in Żyluk et al. (2016), some participants argued that spiciness could be objectively measured, for example, using the Scoville scale (Scoville, 1912). Others argued that the spiciness could be derived from the recipe, in a way that stew containing spicy ingredients such as chilli or pepper must be spicy and the person claiming it’s not can’t be right. Therefore this item for those participants could not have been placed in the personal taste domain. To obtain the item appropriate for this domain, “spicy” was replaced with “tasty,” more subjective.

Another item turned out to be highly knowledge-dependent. Included in the domain of truth about the physical world (*Robin believes one mathematician’s proof of the math formula is right and Chris believes another mathematician’s proof of the math formula is right*) puzzled participants who incorrectly represented what proof is. Some participants in their interviews expressed beliefs that proofs may be incorrect, which suggests that they do not know the definition of mathematical proof. In SEUA-A (the full-abstract version of SEUA), this item was replaced by *Robert thinks that a certain way of calculating the area of a triangle is correct. Chris believes another way of calculating the area of a triangle is correct*, thus providing participants with items of similar content and structure, but with the notion of the way of calculating the area of a triangle familiar even to graduates of primary education.

Overall, 12 items out of 25 underwent change procedures: 2 due to the ambiguity or controversies regarding the content and 10 due to transformations from concrete to abstract. Table 3 presents an exhaustive list of test items in the SEUA version and their replacements developed in the described process, therefore including complete SEUA-A tool. Test items written in a regular font remain unchanged: items italicised underwent the concrete-abstract transformation, items written in bold were changed due to ambiguity, and item 22 was both made abstract and changed its contents.

**Introducing Internet-based version.** As we stressed before, the goal of our research was to create a full-abstract version of SEUA (SEUA-A) and compare it with regular SEUA (used in the same testing conditions as SEUA-A).

As noted in Żyluk et al. (2016), the main disadvantage of SEUA was the difficulty to effectively conduct studies on a bigger sample. This was because administering the questionnaire in the interview setting was time-consuming (up to 40 minutes/participant) and required the interviewer to work with one participant at a time. Taking this into account, the paper-pencil version of SEUA was used in research by Żyluk et al. (2017); however, it performed worse in terms of reliability than the interview version, despite providing precise instructions and stressing the correct ways of responding to test items. As discussed above, this might have been an effect of the abstract-concrete issue, and we hoped that this would be mitigated by the changes, we introduced to the tool.

Since a significant number of studies have shown that Internet-based and paper-pencil versions of questionnaires exhibit similar characteristics (De Beuckelaer & Lievens, 2009; Hedman et al., 2010; Ramsey et al., 2016; Ritter et al., 2004; Weigold et al., 2013) we opted for administering the test via the Internet. This approach has several significant advantages: it does not require the presence of a researcher, allows participants to fill the questionnaire in the most convenient time and place, reduces the costs (no need for printing, transferring the answers from paper to spreadsheet or finding a suitable venue for performing a study) and enables the use of the snowball strategy in collecting data. Internet-based testing poses some challenges, too, including less control over the environment, risks of software and hardware malfunction, fatigue from prolonged screen time (Noyes & Garland, 2008), lower response rate (Kongsved et al., 2007),
Table 3. SEUA and SEUA-A Items’ Comparative Overview.

| No | Domain       | Type       | SEUA                                                                 | SEUA-A                                                                 |
|----|--------------|------------|----------------------------------------------------------------------|----------------------------------------------------------------------|
| 1  | Esthetics    | Abstract   | Robin thinks the first piece of music they listen is to be better.  | Robin thinks the first piece of music they listen is to be better.  |
|    |              |            | Chris thinks the second piece of music they listen is to be better. | Chris thinks the second piece of music they listen is to be better. |
| 2  | The truth about the social world | Abstract   | Robin thinks one book’s explanation of why the Crimean wars began is right. | Robin thinks one book’s explanation of why the Crimean wars began is right. |
|    |              |            | Chris thinks another book’s explanation of why the Crimean wars began is right. | Chris thinks another book’s explanation of why the Crimean wars began is right. |
| 3  | Esthetics    | Concrete   | Robin thinks that porcelain figures are the most beautiful.        | Robin thinks that porcelain figures are the most beautiful.        |
|    |              |            | Chris thinks that glass figures are the most beautiful.            | Chris thinks that glass figures are the most beautiful.            |
| 4  | Moral values | Concrete   | Robin thinks that cheating on a partner is inexcusable.            | Robin has a certain opinion on the acceptability of lying.          |
|    |              |            | Chris thinks that one can forgive a partner who was cheating in certain cases. | Chris has a different view on the acceptability of lying.          |
| 5  | The truth about the physical world | Abstract   | Robin agrees with one book’s explanation of the origin of life on Earth. | 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. | Chris agrees with another book’s explanation of the origin of life on Earth. |
| 6  | Personal taste | Concrete | Robin says that too goes best with breakfast.                     | Robin says that a beverage goes best with breakfast.                |
|    |              |            | Chris says coffee goes best with breakfast.                        | Chris says coffee goes best with breakfast.                        |
| 7  | Moral values | Concrete   | Robin thinks lying is wrong.                                       | Robin has a certain opinion on the acceptability of lying.          |
|    |              |            | Chris thinks lying is permissible in certain situations.           | Chris has a different view on the acceptability of lying.          |
| 8  | Personal taste | Abstract | Robin says the stew is spicy.                                      | Robin says the stew is tasty.                                       |
|    |              |            | Chris says the stew is not spicy at all.                          | Chris says the stew is not tasty at all.                           |
| 9  | Esthetics    | Abstract   | Robin thinks the first painting they look at is better.           | Chris thinks the first painting they look at is better.            |
|    |              |            | Chris thinks the second painting they look at is better.          | Chris thinks the second painting they look at is better.          |
| 10 | Personal taste | Concrete | Robin thinks weddings should be held in the afternoon.             | Robin thinks that a certain time of the day is more appropriate for weddings. |
|    |              |            | Chris thinks weddings should be held in the evening.               | Chris thinks that another time of the day is more appropriate for weddings. |
| 11 | The truth about the physical world | Abstract   | Robin believes one mathematician’s proof of the math formula is right. | Robin believes that a certain way of calculating a triangle’s area is correct. |
|    |              |            | Chris believes another mathematician’s proof of the math formula is right. | Chris believes another mathematician’s proof of the math formula is right. |
| 12 | Moral values | Concrete   | Robin believes that the well-being of an individual is more important than the common good. | Robin believes that the well-being of another part of society is more important than the well-being of others. |
|    |              |            | Chris believes that the common good is more important than the well-being of an individual. | Chris believes that the well-being of another part of society is more important than the well-being of others. |
| 13 | Moral values | Concrete   | Robin thinks the government should limit the number of children families are allowed to have to keep the population from getting too big. | Robin thinks that certain pro-family policy is right. |
|    |              |            | Chris thinks families should have as many children as they choose. | Robin believes one book’s explanation of what atoms are made up of. |
| 14 | The truth about the physical world | Abstract   | Robin believes one book’s explanation of what atoms are made up of. | 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. | Chris believes another book’s explanation of what atoms are made up of. |
| 15 | The truth about the social world | Abstract   | Robin has one view of why criminals keep going back to crime.      | 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. | Chris has a different view of why criminals keep going back to crime. |
| 16 | Personal taste | Concrete | Robin says warm summer days are the nicest.                        | Robin says that another season is the nicest.                     |
|    |              |            | Chris says cool autumn days are the nicest.                        | Chris says that another season is the nicest.                     |
| 17 | The truth about the physical world | Abstract   | Robin believes one book’s explanation of how the brain works.      | Robin believes one book’s explanation of how the brain works.      |
|    |              |            | Chris believes another book’s explanation of how the brain works. | Chris believes another book’s explanation of how the brain works. |
| 18 | Esthetics    | Abstract   | Robin thinks the second book they both read is better.            | Robin thinks the second book they both read is better.            |
|    |              |            | Chris thinks the second book they both read is better.            | Chris thinks the second book they both read is better.            |
| 19 | The truth about the social world | Abstract   | Robin has one view on the causes of the increased number of divorces. | Robin has one view on the causes of the increased number of divorces. |
|    |              |            | Chris has a different view on the causes of the increased number of divorces. | Chris has a different view on the causes of the increased number of divorces. |
| 20 | Esthetics    | Abstract   | Robin thinks the first car they saw was more beautiful.           | Robin thinks the first car they saw was more beautiful.           |
|    |              |            | Chris thinks the second car they saw was more beautiful.          | Chris thinks the second car they saw was more beautiful.          |
| 21 | The truth about the social world | Abstract   | Robin has one view on the causes of unemployment.                 | Robin has one view on the causes of unemployment.                 |
|    |              |            | Chris has a different view on the causes of unemployment.         | Chris has a different view on the causes of unemployment.         |
| 22 | Moral values | Concrete   | Robin thinks people should take responsibility for themselves.    | Robin has a certain opinion on the use of animal products.         |
|    |              |            | Chris thinks people should work together to take care of each other. | Chris has a different opinion on the use of animal products.     |
| 23 | The truth about the social world | Abstract   | Robin agrees with one book’s explanation of how children learn a language. | Robin agrees with one book’s explanation of how children learn a language. |
|    |              |            | Chris agrees with another book’s explanation of how children learn a language. | Chris agrees with another book’s explanation of how children learn a language. |
| 24 | Personal taste | Concrete | Robin thinks that a soft mattress is the most comfortable to sleep on. | Robin thinks that one kind of mattress is the most comfortable to sleep on. |
|    |              |            | Chris thinks that a firm mattress is the most comfortable to sleep on. | Chris thinks that another kind of mattress is the most comfortable to sleep on. |
| 25 | The truth about the physical world | Abstract   | Robin agrees with some explanation of the pancreatic cancer causes. | Robin agrees with a certain explanation of the pancreatic cancer causes. |
|    |              |            | Chris agrees with a different explanation of the pancreatic cancer causes. | Chris agrees with a different explanation of the pancreatic cancer causes. |

or inappropriateness for more sophisticated studies of personal preferences (Windle & Rolfe, 2011); we are convinced, however, that those limitations are not very likely to threaten the performance of SEUA, as the test is short, simple, and administered within reliable system.

Introducing the Internet version of the tool also allowed some functional changes in administering SEUA. Following concerns that arose in previous studies that the participants might be tempted to adjust their answers to follow their earlier choices in a given domain, in the Internet-based SEUA-A version, we disabled the possibility to revisit previous test items. This was to reduce the chance that participants match test items into sets consistent with domains of the SEUA, given that this would require extensive use of working
memory. We did not control how the participants displayed the supporting schema of answering, so they had a free choice of having it on display all the time (e.g., on-screen shared between the window with a questionnaire form) or revisiting it when needed (e.g., in a different tab in the web browser). We also forced the participants to make a choice for any given item, not allowing them to proceed without selecting the answer.

Links to both questionnaires were distributed among students and young graduates to obtain a sample that was comparable regarding the age and education (and, thus, according to Kuhn et al., 2000, to the levels of their epistemological understanding) with samples from previous studies that included earlier versions of our questionnaires measuring epistemological understanding (Żyluk et al., 2016, 2017). The primary goal of this strategy was to enable easier comparisons between results obtained in the context of using our tools. In the research described in Żyluk et al. (2016) the sample consisted of 40 adults with ages ranging from 19 to 35, mostly \(n = 31\) students and in the research described in the text by Żyluk et al. (2017) the sample included 47 students with ages ranging from 20 to 25. It should be noted, however, that the original study conducted by Kuhn et al. (2000) included the sample that was substantially different in terms of the homogeneity from ours—the sample in the study conducted by Kuhn et al. (2000) consisted of seven groups of participants varying in age, life, and educational experience (5th grade, 8th grade, 12th grade, undergraduate, community college, professional, and expert), as the main objective of that study was to assess if epistemological understanding develops in the predicted order across judgment domains.

The questionnaires were distributed using a social media channel (for its further specification, see subsection 3.3). Given our strategies for acquiring respondents, we can conclude that we utilized the variant of the opportunity sampling method in our research. To ensure that one person did not fill both SEUA and SEUA-A, the disclaimer asking individuals to refrain from participating in the study if they had already filled any previous version was also placed on the introduction page of the Internet questionnaire.

**Participants**

Overall, 108 people participated in the presented study, with 58 filling the SEUA questionnaire, and 50 filling the SEUA-A questionnaire. The gender proportions were balanced for SEUA (35 female, 23 male) and not balanced for SEUA-A (39 females and 11 males). We aimed at collecting a sample that ranged from 20 to 30 years old. The mean age for SEUA-A participants was 24.70 \(SD = 2.38\) with the youngest participant aged 20 and the oldest 30. For SEUA, the mean age was 22.91 \(SD = 2.91\), with participant ages ranging from 19 to 30. Participants were also asked about their educational background (field of studies and if they graduated; if they did not graduate they needed to specify their year of studies). Participants declared very diverse academic background. Amongst declared fields of study, none was dominant or particularly frequently indicated: 12 participants for SEUA and 19 participants for SEUA-A had already graduated (at Bachelor/Master level).

**Procedure**

The study was conducted via the Internet. A short survey with demographic questions accompanied the questionnaire (see above), and space for questions/remarks was administered using Google Forms. The information about the study with a link to the online questionnaire was distributed in social media, both in groups that at least partially included people of target age and education (large local Facebook groups, student Facebook groups) and on the authors’ private social media profile(s). In balanced proportions, posts inviting participation in the study included one of two links: one leading to the SEUA or one leading to SEUA-A. The instruction, demographic questions and visual layout of the questionnaire did not differ. Instructions informed participants about the purpose of the study, presented a schema for answering questions, and assured the participants that the results would be used for scientific purposes only, in an anonymized way. We also stressed that the participants should try not to let their personal views on given topics influence the answers and that there were no right or wrong answers.

To calculate the scores for SEUA and SEUA-A in the online versions, we used the same procedure, briefly described in Section 2.1. Every A answer was scored as 1 point, M, 2 points, and E, 3 points. The summary score for every domain was calculated for each participant. This quantitative approach allowed to run the reliability analysis. It is worth noting, however, that SEUA-A, like its predecessor, also maintains the ability for scores to be analyzed qualitatively and to determine the participants’ epistemological understanding profile (for more about qualitative analysis and profiles, see Section 2.1.), even though it was not used in this study.

Table 4 presents an overview of all discussed versions of the tool, from the original work of Kuhn et al. (2000) to the version which is a topic of this article. To sum up, the original version was administered in a paper-pencil setting and included 15 items—three in each domain, and had no quantitative scoring method. SEUA was used three times with different administration modes: interview, paper-pencil and Internet; it included 25 items—five in each domain, and had a scoring method with possible scores ranging from 25 to 75. SEUA-A was used only in Internet settings, consisted of 25 qualitatively scored items and was the only version where no concrete items were present. Please note that since the modifications that led to SEUA didn’t include changing the original items, the 15 items that constituted the original version
were nested in SEUA. Therefore, it was possible to calculate reliability scores for this set of items in the studies described in Żyluk et al. (2016, 2017) and here.

**Results**

Data collected in the study were processed in IBM SPSS Statistics v. 24.

**Descriptive Statistics**

Descriptive statistics for participants scores for the Internet-based SEUA and SEUA-A are displayed in Table 5. Minimal and maximal values in the table correspond to the actual values obtained by participants. The normality of the scores’ distribution was assessed using the Kolmogorov-Smirnov test. We did not observe any missing answers, as we forced the participants to choose their options in order to proceed.

**Reliability**

To assess the internal consistency of the tool, Cronbach’s alpha coefficient was used. Results of the analysis are presented in Table 6. To put the tool’s reliability in perspective and to compare it with previous versions, we have included Cronbach’s alpha coefficients obtained in previous research in this table: SEUA in an interview setting and SEUA in a paper-pencil setting.

Analysis of Cronbach’s alpha coefficient for SEUA administered via the Internet shows that the reliability of that tool decreases when changing the mode from interview to the Internet-based. In the already sub-reliable domain of personal taste (Żyluk et al., 2016), Internet administration of SEUA resulted in Cronbach’s alpha dropping to alpha = .100. Internal consistency estimation did not suffer from such a drastic decrease for the other domains but was lower nonetheless. The only change in Cronbach’s alpha in favour of the Internet-based SEUA can be seen in the domain of truth about the social world, with the negligible change from alpha = .775 in the interview setting to alpha = .806 for the Internet-based SEUA.

On the other hand, achieved levels of Cronbach’s alpha for the SEUA-A version show that applied changes improved the tool in terms of internal consistency. The most noticeable change in a value can be seen in the case of the personal taste domain. In the authors’ previous research Cronbach’s

---

**Table 4. Instruments’ Overview.**

|                | Original version | SEUA | SEUA | SEUA | SEUA-A |
|----------------|------------------|------|------|------|--------|
| Reference      | Kuhn et al. (2000) | Żyluk et al. (2016) | Żyluk et al. (2017) | Present paper | Present paper |
| Setting        | Paper-pencil | Interview | Paper-pencil | Internet | Internet |
| Items          | 15 | 25 (15 for nested original items) | | 25 |
| Minimal score (theoretical) | — | 25 | 75 (45 for nested original items) | — |
| Maximal score (theoretical) | — | 75 | — |
| Concrete items | Yes | No |
| Notes          | No quantitative scoring | Included (nested) short, 15-item original version | — |
| Sample size in the study | 129 (7 groups) | 40 | 47 | 58 | 50 |

**Table 5. Descriptive Statistics for Internet-Based SEUA and SEUA-A Scores.**

|                | Min | Max | M   | Me   | SD  | Z  | p-Value |
|----------------|-----|-----|-----|------|-----|----|---------|
| SEUA-personal taste | 7.00 | 12.00 | 10.12 | 10.00 | 0.987 | 0.279 | <.001 |
| SEUA-esthetics    | 8.00 | 14.00 | 10.50 | 10.00 | 1.158 | 0.374 | <.001 |
| SEUA-values       | 8.00 | 14.00 | 10.50 | 10.00 | 1.158 | 0.374 | <.001 |
| SEUA-truth: social | 7.00 | 15.00 | 12.72 | 13.00 | 2.22 | 0.193 | <.001 |
| SEUA-truth: physical | 7.00 | 15.00 | 12.72 | 13.00 | 2.22 | 0.193 | <.001 |
| SEUA-total        | 39.00 | 67.00 | 56.66 | 58.00 | 6.089 | 0.105 | .172 |
| SEUA-A-personal taste | 5.00 | 15.00 | 10.32 | 10.00 | 1.50 | 0.356 | <.001 |
| SEUA-A-esthetics  | 5.00 | 15.00 | 10.30 | 10.00 | 1.57 | 0.364 | <.001 |
| SEUA-A-values     | 5.00 | 15.00 | 10.30 | 10.00 | 1.57 | 0.364 | <.001 |
| SEUA-A-truth: social | 5.00 | 15.00 | 13.10 | 14.00 | 2.11 | 0.185 | <.001 |
| SEUA-A-truth: physical | 5.00 | 15.00 | 13.10 | 14.00 | 2.11 | 0.185 | <.001 |
| SEUA-A-total      | 25.00 | 75.00 | 57.12 | 58.00 | 7.342 | 0.137 | .018 |

Note. Min = minimum; Max = maximum; M = mean; Me = median; SD = standard deviation; Z = KS test statistics; p = significance level of a Z.
alpha for personal taste domain ranged from .100 to .439, while in the SEUA-A version, alpha = .825 was obtained. Only in the domain of the truth about the social world, the reliability measure was marginally lower in SEUA-A (alpha = .795) compared to SEUA administered via the Internet (alpha = .806). In the case of the remaining domains, a noticeable improvement could be seen compared to SEUA-Internet-based and SEUA in an interview setting. The lowest level of Cronbach’s alpha was obtained for the values domain (alpha = .736). However, in comparison to reliability measures for the values domain in other versions and settings, SEUA-A was still superior.

Additional Qualitative Data

In the Internet-based forms, we provided space for questions and remarks, and we briefly analyzed the feedback that participants offered. For SEUA, seven individuals indicated problems understanding the questions or expressed some doubts regarding how they were supposed to answer them. For SEUA-A, only one person reported similar difficulties. We are aware that this data cannot be conclusive and that we would be more likely to hear the full spectrum of participants’ remarks in the interview. Still, it may be seen as an indication of the greater intelligibility of the task in SEUA-A.

Discussion

Our goal while developing SEUA-A was to produce a tool for measuring levels of epistemological understanding that demonstrates a satisfactory level of reliability and is easy to administer. Earlier versions suffered from either a lack of quantitative scoring (original tool), complicated administration (SEUA-interview) or low reliability for certain domains (SEUA-Internet, SEUA paper-pencil, and SEUA-interview). Introduced changes mitigated these main problems: SEUA-A is a satisfactorily reliable tool that can be administered without the need for researchers’ supervision over each participant.

The main focus of this improvement process was the abstract-concrete division in test items. We succeeded in replacing each concrete item with a closely corresponding abstract version, which resulted in a more homogenous test form. When participating in the study which included SEUA-A, participants no longer had to switch between those two types of items. The homogeneously abstract version produced better reliability scores—and they probably arise not only from including more homogenous items but also from a better understanding of the task by participants and a lower level of influence of their personal opinions.

The administration of the test was also significantly altered. While the interviews conducted for developing SEUA provided invaluable information regarding participants’ thought processes, difficulties, and questions, they were not a method appropriate for use in more extensive studies. Simply translating the tool to paper-pencil or the Internet version was insufficient for obtaining satisfactory outcomes. Since the results improved after removing ambiguity arising from the presence of concrete items, a plausible explanation of the worse performance of the tool in the paper-pencil version is the lack of researchers’ help in the process. While the completion times were not measured in the study described in this paper, each item’s lack of explanatory component(s) should significantly reduce testing time.

We are aware that other tools for measuring the levels of epistemological understanding were developed before (see, e.g., Braten, 2010; Muis et al., 2006; Sandoval et al., 2016 for the overview of the existing tools). However, SEUA-A has a significant advantage of being an easy to administer, satisfactorily reliable tool with a low risk of results being tainted with, for example, personal preferences of the participants. Another reason to use SEUA-A is to use qualitative profiles to understand better individual trajectory in the development of epistemological understanding and quantitative scores for statistical analysis and reliability testing.

The empirical results reported herein should be considered in light of some limitations. The primary limitation to the generalization of these results is the sample size—the number of participants can be seen as relatively limited to end up with any firm conclusions. To confirm our results, in the next step, the study should be repeated on a larger and preferably more heterogeneous (e.g., in terms of age and education) sample, to allow not only the general analysis but also to perform inter-group comparisons (perhaps in a similar way as was done in the research by Kuhn et al. (2000)). The second limitation concerns the nonrandom assignment of participants to the version of the instrument. Redirecting to a specific instrument version should be more automated in any possible future research involving both

| Domain               | SEUA-interview | SEUA-paper-pencil | SEUA-Internet | SEUA-A-Internet |
|----------------------|----------------|--------------------|---------------|-----------------|
| Personal taste       | 0.439          | 0.436              | 0.100         | 0.825           |
| Esthetics            | 0.852          | 0.767              | 0.657         | 0.858           |
| Values               | 0.662          | 0.647              | 0.604         | 0.736           |
| Truth: social world  | 0.775          | 0.721              | 0.806         | 0.795           |
| Truth: physical world| 0.745          | 0.670              | 0.677         | 0.792           |
tools. The strategy we employed in our study was not optimal. We assumed that publishing different versions of the tool in different groups and adding the disclaimer asking individuals to refrain from participating in the study if they had already filled the second version of the tool would suffice. A certain limitation is also our sampling method—recruiting participants via social media by posting in groups that at least partially included people of target age and education plus on the authors’ private profile(s). Even though this method of reaching the respondents was dictated by the nature of our target group (students and young graduates), it was definitely possible to diversify these methods even on the Internet itself, for example, publishing links on students’ forums.

Additionally, one of the weakest points of the study is the lack of extensive validity analyzes. The opinion of competent judges supported the SEUA-A’s final content (on two levels: firstly, we rely on the opinion of the judges while creating SEUA—that is, while adding new items to the original tool by Kuhn et al. (2000) (see a paper by Żyluk et al., 2016), and secondly, in the present research, when we were transforming concrete items from SEUA into abstract ones, which resulted in having SEUA-A) and this tool exhibits satisfactory level of reliability within judgment domains. However, these processes and results seem not to suffice to conclude that SEUA-A is indisputably a valid tool. In a possible future study, one of the methods of performing the reliability analysis would be to group comparisons, that is, checking whether differences between groups are consistent with the assumptions suggesting that views on the nature of knowledge and knowing of an individual can be influenced by one’s education, life experience and age (Hofer & Pintrich, 1997; Kuhn et al., 2000). Another method of reliability analysis would be checking whether obtained individual letter profiles’ patterns are consistent with the hypothesized (Kuhn et al., 2000) order of transitions between levels of epistemological understanding (see section 2. for the description of the assumed sequence of these transitions). In our previous study in an interview setting (Żyluk et al., 2016), we performed profiles’ analysis and obtained results mostly consistent with the hypothesized order of transitions between levels of epistemological understanding (“mostly”—but not entirely, as for some domains the number of participants exhibiting inconsistent patterns was up to 27.5%). Another source of validity confirmation would be correlational analysis. The existence of the correlation between SEUA-A results and argumentation skills (measured, e.g., similarly as in the research by Mason & Scirica, 2006 or by Nussbaum et al., 2008) or results obtained using other methods for measuring the development of personal epistemology (such as The Epistemic Beliefs Inventory (EBI) by Schraw et al., 2002 or Epistemic Thinking Assessment (ETA) by Barzilai & Weinstock, 2015) would help provide external validity for the SEUA-A instrument.

When it comes to recommendations which version is most optimal for use in further research, the response varies. The SEUA-A version can be recommended to use in paper-pencil and Internet-based studies and test batteries, without the need for individual researcher-participant interactions in most cases. While preserving high reliability, it can be applied quickly to multiple participants at a time at a fraction of the cost of conducting interviews. It remains an open question if items from SEUA-A are suitable for an interview setting. A one-on-one interactive approach to using this test during interviews can be further tested, and it is expected to perform no worse than SEUA. As for further research, there seems to be little room for improvement when Kuhn’s et al. (2000) dimensions are taken as a basis of tool construction. Potentially, after conducting interviews with items from SEUA-A, one can get a better insight into the response process, which later can serve as an indication for future improvement possibilities (as this was the case with the previous SEUA version). Furthermore, as SEUA-A allows for time-efficient, satisfactorily reliable testing of personal epistemology within an adapted theoretical framework, research on epistemological understanding can benefit from developing tools that explore more diverse theoretical approaches to the topic.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research reported in this paper were supported by the National Science Centre, Poland (DEC-2013/10/E/HS1/00172).

ORCID iD

Mariusz Urbański https://orcid.org/0000-0002-8682-5307

References

Ahola, S. (2009). Measurement issues in studying personal epistemology. *Psychology & Society, 2*(2), 184–191.
Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemporary Educational Psychology, 42*, 141–158.
Barzilai, S., & Zohar, A. (2009). The role of epistemic thinking in online learning. In Y. Eshet-Alkalai, A. Caspi, S. Eden, N. Geri, & Y. Yair (Eds.), *Learning in the technological era: Proceedings of the 4th Chais conference on instructional technologies research* (pp. 29–33). The Open University of Israel.
Braten, I. (2010). Personal epistemology in education: Concepts, issues, and implications. In E. Baker, B. McGaw, & P. Peterson (Eds.), *International encyclopedia of education* (pp. 211–217). Elsevier.
Christodoulou, A., Osborne, J., Richardson, K., Howell-Richardson, C., & Simon, S. (2010, March 21–24). A study of student beliefs about the epistemology of science and their relationship with students’ personal epistemologies [Conference session]. The NARST conference, Philadelphia, PA.

De Beuckelaer, A., & Lievens, F. (2009). Measurement equivalence of paper-and-pencil and Internet organizational surveys: A large scale examination in 16 countries. Applied Psychology, 58(2), 336–361.

Hedman, E., Ljötsson, B., Ruck, C., Furmark, T., Carlbring, P., Lindefors, N., & Andersson, G. (2010). Internet administration of self-report measures commonly used in research on social anxiety disorder: A psychometric evaluation. Computers in Human Behavior, 26(4), 736–740.

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. Educational Psychology Review, 13(4), 353–383.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological beliefs: theories about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88–140.

Holma, K., & Hyytinen, H. (2015). The philosophy of personal epistemology. Theory and Research in Education, 13(3), 334–350.

Kongsved, S. M., Basnóv, M., Holm-Christensen, K., & Hjollund, N. H. (2007). Response rate and completeness of questionnairenaires: A randomized study of Internet versus paper-and-pencil versions. Journal of Medical Internet Research, 9(3), e25.

Kuhn, D. (1991). The skills of argument. Cambridge University Press.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. Cognitive Development, 15(3), 309–328.

Labbas, R. (2013). Epistemology in education: Epistemological development trajectory. Journal of International Education & Leadership, 3(2), 1–10.

Madjar, N., Weinstock, M., & Kaplan, A. (2017). Epistemic beliefs and achievement goal orientations: Relations between constructs versus personal profiles. The Journal of Educational Research, 110(1), 32–49.

Mason, L., Boldrin, A., & Zurlo, G. (2006). Epistemological understanding in different judgment domains: Relationships with gender, grade level, and curriculum. International Journal of Educational Research, 45(1), 43–56.

Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in interpreting a controversy and in topic-specific belief change. Contemporary Educational Psychology, 29(2), 103–128.

Mason, L., & Scirica, F. (2006). Prediction of students’ argumentation skills about controversial topics by epistemological understanding. Learning and Instruction, 16, 492–509.

Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. Educational Psychology Review, 18(1), 3–54.

Noroozi, O. (2016). Considering students’ epistemic beliefs to facilitate their argumentative discourse and attitudinal change with a digital dialogue game. Innovations in Education and Teaching International, 53(3), 1–9.

Noyes, J. M., & Garland, K. J. (2008). Computer-vs. paper-based tasks: Are they equivalent? Ergonomics, 51(9), 1352–1375.

Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. International Journal of Science Education, 30, 1977–1999.

Perry, W. G., Jr. (1970). Forms of intellectual and ethical development in the college years: A scheme. Holt.

Ramsey, S. R., Thompson, K. L., McKenzie, M., & Rosenbaum, A. (2016). Psychological research in the internet age: The quality of web-based data. Computers in Human Behavior, 58, 354–360.

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: A randomized comparison. Journal of Medical Internet Research, 6(3), e30.

Sandoval, W. A., Greene, J. A., & Braten, I. (2016). Understanding and promoting thinking about knowledge: Origins, issues, and future directions of research on epistemic cognition. Review of Research in Education, 40(1), 457–496.

Schraw, G., Bendixen, L. D., & Dunkle, M. E. (2002). Development and validation of the epistemic belief inventory. In B. K. Hofer & P. R. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp. 261–275). Erlbaum.

Scoville, W. L. (1912). Note on capsicums. Journal of Pharmaceutical Sciences, 7(5), 453–454.

Tabak, I., & Weinstock, M. (2008). A sociocultural exploration of epistemological beliefs. In M. Swe Khine (Ed.), Knowing, knowledge and beliefs (pp. 177–195). Springer.

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. Psychological Methods, 75(1), 33.

Windle, J., & Rolfe, J. (2011). Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. Economic Analysis and Policy, 47(1), 83.

Żyluk, N., Karpe, K., Michta, M., Potok, W., Paluszkiwicz, K., & Urbański, M. (2016). Assessing levels of epistemological understanding: The Standardized epistemological understanding assessment (SEUA). Topoi, 37, 1–13.

Żyluk, N., Michta, M., & Urbański, M. (2017). Yet another shade of deduction. On measuring deductive flexibility and how it may relate to other cognitive abilities. Logic and Logical Philosophy. Advance online publication. https://doi.org/10.12775/LLP.2017.015