CRITICAL THINKING AS AN IMPACT FACTOR ON STATISTICAL LITERACY – THEORETICAL FRAMEWORKS AND RESULTS FROM AN INTERVIEW STUDY

Einav Aizikovitsh-Udi[1] and Sebastian Kuntze[2]
[1]Beit Berl Academic College, Israel, [2]Ludwigsburg University of Education, Germany
eia994@mail.harvard.edu, kuntze@ph-ludwigsburg.de

The theoretical frameworks of Critical Thinking (CT) and Statistical Thinking (ST) suggest an overlap – however, the quality of the connectedness of CT and ST has still not been described empirically in a satisfactory way. As elements of ST are a key prerequisite of statistical literacy, CT impacts on statistical literacy as well. This study hence focuses on the role of CT in the process of solving problems which require statistical literacy. A case analysis based on interview data provides insight into thinking processes and affords focusing such connections between CT and ST. The results support the hypothesis that thinking skills in both areas are interdependent and help to describe key intersection areas from a theoretical point of view: For instance, the interviewees' use of strategies of evaluating claims has a high explanatory power and provides a combined framework from the CT and ST perspectives.

INTRODUCTION

Both critical thinking (CT) and statistical literacy (SL) refer to areas of thinking skills and dispositions which are necessary for social participation of responsible citizens in democratic societies. Whereas CT focuses on rather general thinking skills and strategies which are considered useful almost regardless of specific content domains, SL concentrates on learners' strategies of dealing with data and statistical representations or models. Even if relationships between CT and SL may appear obvious from the theoretical level, exact descriptions and empirical evidence of how CT can contribute to statistical thinking (ST) processes is scarce. Consequently, the project CCTST ("Connections between Critical Thinking and Statistical Thinking") aims at exploring how and to what extent components of CT are requirements for ST and how CT may also impact on SL. In this paper, we present evidence from a series of interviews which suggests that a combined perspective on ST and CT can explain thinking processes as well as difficulties thinkers encounter when having to solve problems referring to statistical contexts. In particular, a focus on strategies of evaluating claims affords distinguishing between "CT within ST" and "CT enriching ST".

In the following first section, we will briefly introduce the theoretical background the study refers to. The second section will summarize the research questions. In the third section we will describe the methods we used in this interview study. Results are given in the fourth section; these results will be discussed in the concluding fifth section.

THEORETICAL BACKGROUND

According to a well-known definition by Wallman (1993), SL is "the ability to understand and critically evaluate statistical results that permeate our daily lives—coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions". This definition suggests that for SL, abilities of using of statistical models and representations are in the centre, and that SL is connected with the learners' achievement in this statistics-related area of expertise. We hence consider SL as a competency construct (cf. Watson & Callingham, 2003; Kuntze, Lindmeier, & Reiss, 2008; Kuntze, Engel, Martignon & Gundlach, 2010), and we assert that this competency encompasses ST skills. Requirements connected to SL can be described according to the aspects of *dealing with statistical variation* (e.g. Watson & Callingham, 2003), *manipulating data by reduction* (e.g. Kröpfl, Peschek, & Schneider, 2000), and *data-related reading* (Curcio, 1987), which are crucial aspects of many problems in statistical contexts. These aspects have been integrated in a theory-based hierarchical competency model which maps learners' performance in the domain of "using models and representations in statistical contexts" (see e.g. Kuntze, Linmeier & Reiss, 2008, and Kuntze, et al., 2010, for details). The more statistically literate learners are, the more they are able to draw on conceptual knowledge and strategies related to the three above-mentioned aspects (cf. also Shaughnessy, 2007).

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

Even more than it is the case for SL, there is a broad variety of different theoretical approaches to CT (e.g. Dewey, 1933, p. 118; Ennis, 1987; 2002; Lipman, 1991; McPeck, 1981). McPeck (1981) for example describes CT as “skills and dispositions” necessary to “appropriately use reflective skepticism”. Lipman (1991) sees CT as “thinking that enables judgment, is based on criteria, corrects itself, and is context-sensitive.” In the project CCTST, we refer to the framework developed by Ennis (1987; 1989; 2002; Ennis & Millman, 2005), which consists of sub-constructs that have also been implemented in a corresponding test instrument (Ennis & Millman, 2005). In Ennis’ (1987) pragmatic words, CT is "reasonable, reflective thinking that is focused on deciding what to believe and do". Among the mentioned CT sub-constructs are abilities such as analyzing arguments, raising clarification questions, evaluating the source’s reliability or inductive inference, for instance. Even though a careful examination of the corresponding constructs is necessary, already the mentioned aspects of CT suggest that there is a relationship between CT and ST.

Despite existing catalogues of criteria (Ennis, 1987; Wild and Pfannkuch, 1999) both the frameworks of CT and ST remain rather general with some of the notions used and with respect of connections with each other: In particular, there is no consistent theoretical framework though which both ST and CT can be accessed.

However, one core aspect which is in the foreground of both CT and ST consists in ways of dealing with assumptions, hypotheses or statements. For this particular context, dealing with hypotheses in the lens of the scientific reasoning approach (Kuhn, 1989; Kuhn, Amsel & O’Loughlin, 1988; Dunbahr & Klahr, 1989; Bullock & Ziegler, 1994) provides a framework which affords describing elements of CT and ST in a joint perspective. We hence used this framework for a corresponding analysis in this study, focusing simultaneously on CT and ST (cf. Kuntze, Aizikovitsh-Udi & Clarke, 2013). Indeed, there are strategies of e.g. actively questioning claims which are relatively independent from statistical models – however, there are also content-specific strategies, i.e. checking how “statistically stable” a claim derived from data may be or whether an interpretation of data is unaware of phenomena connected with data reduction.

RESEARCH INTEREST

The theoretical background sketched in the previous section highlights the need of research into connections between CT and ST. The study hence focuses on the following research questions:

(1) How can the thinking processes of the interviewees be interpreted from a CT perspective, from a ST perspective and from a combined CT and ST perspective?
(2) What insight about the interrelatedness of CT and ST can be obtained from a focus on the use of strategies of evaluating claims integrated into these analysis steps?

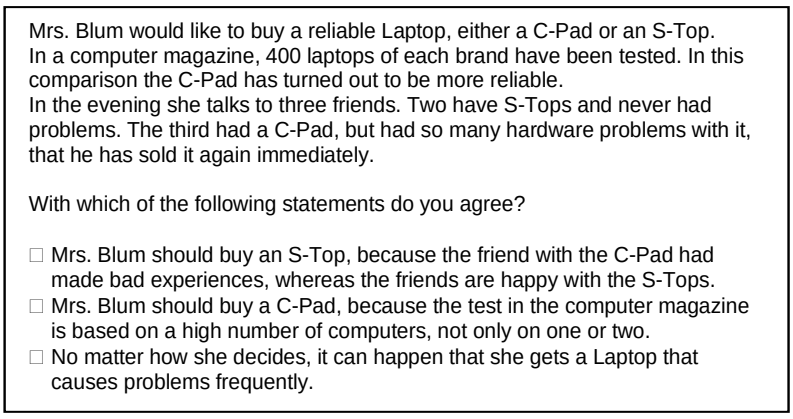
RESEARCH DESIGN AND METHODS

As a first step in the work of the project CCTST, qualitative data was collected in order to explore thinking processes of adults who are confronted with problems which require SL. The problems were derived from the studies by Kuntze, Lindmeier and Reiss (2008) as well as Kuntze et al. (2010). In further steps of the project work, we plan also to include students.

The interview data was gathered using the thinking-aloud method; the role of the interviewer was to encourage the interviewees to verbalize their thoughts and their reasoning. The data can give insight into cases of thinking processes and hence help to generate hypotheses about the interplay of thinking elements which may be interpreted as evidence of CT or ST. Consequently, in a bottom-up interpretive procedure (cf. Mayring, 2000), the evidence was first analyzed from a CT perspective only, using criteria from the approaches presented above. Separately, another interpretive analysis was done against the background of criteria which are relevant for ST. In a third step, a combined analysis was carried out which concentrated on aspects from both the CT and ST frameworks. According to the thoughts developed in the theoretical background section, the specific focus on strategies of evaluating claims was in the foreground of this third analysis, in particular.

As a consequence of length limits, we will concentrate on one problem the interviewees had to solve: the problem “laptops” (Fig. 2). This problem gives three statements which have to be evaluated. We assert that the given context situation requires the active production of a model (cf. framework developed in Kuntze et al., 2008), which again is likely to encourage dealing with

claims and assumptions. The phenomena we observed for this problem were typical also for other problems. In particular, the findings presented in the following section were not limited to this specific “laptop” problem.


Mrs. Blum would like to buy a reliable Laptop, either a C-Pad or an S-Top.
In a computer magazine, 400 laptops of each brand have been tested. In this comparison the C-Pad has turned out to be more reliable.
In the evening she talks to three friends. Two have S-Tops and never had problems. The third had a C-Pad, but had so many hardware problems with it, that he has sold it again immediately.

With which of the following statements do you agree?

□ Mrs. Blum should buy an S-Top, because the friend with the C-Pad had made bad experiences, whereas the friends are happy with the S-Tops.
□ Mrs. Blum should buy a C-Pad, because the test in the computer magazine is based on a high number of computers, not only on one or two.
□ No matter how she decides, it can happen that she gets a Laptop that causes problems frequently.

Figure 2: Problem “Laptops”

RESULTS

We will in the following give examples of evidence from the interviews connected to the problem in Figure 2 together with the results of the interpretive analysis. We start with data from the interview with Alex, who is an experienced secondary teacher from Israel. After a short time of thought, Alex answered in the following way:

Alex: Assuming the magazine test was conducted professionally and fairly – I would agree with the last two statements.
Yes, the test of the computer magazine is based on a much higher number of computers and is more reliable then.
The hardware problems the third friend had with his C-Pad should most likely be attributed to the possibility described by the last statement. It is more probable that he just happened to buy an exceptionally bad C-pad unit than the magazine test happened to check an exceptional 400 units sample that turned out the wrong result.
Of course it is highly advisable to also check, if possible, the statistical significance of the magazine test.

Analyzing Alex’ answers from the *ST perspective*, we do not only see a correct choice of the options, but we also observe several facets that highlight a very high level of ST. Firstly, Alex questions the design conditions of the magazine test and points out that quality standards such as professionality and fairness play a role. By these aspects, he might have thought of the question of sampling, together with other quality standards such as validity or objectivity. However, not only these criteria are important to him, but also the criterion of statistical significance is mentioned by Alex in the last sentence of his answer.

Secondly, he uses the number of computers included in the test as an indicator of a higher reliability.

Thirdly, he argues using the notion of probability. In this model, he compares the scenario of drawing a sample of size one of an “exceptionally bad” C-Pad with drawing a sample of 400 exceptionally good C-Pads. Even if this reasoning appears as questionable in a detail consideration (not all 400 tested Laptops need to be ‘exceptional’ for having a ‘wrong’ result of the study, for example), the answer makes use of models specific for ST and may be seen as representative for supportive intuitions that link the sample size to the corresponding probabilities.

An additional observation is that the model mentioning probability can be seen as an attempt to provide a deepened justification for his first claim, that a higher number of computers corresponds to a higher reliability. Alex thus appears to reflect deeply from a ST point of view and he probably aims to provide reasons for his correct answers, using mathematical models. The

presentation form of these reasons however is not formal; the main target of Alex might be to convince himself and/or the interviewer of the answers he gives to the questions.

From the *CT perspective*, Alex not only analyses statements by comparing e.g. their reliability, but he also evaluates the reliability of the source, when discussing methodological standards and properties of the computer magazine study. In particular, the first sentence of Alex' answer could point into the direction of a less ST-like reasoning but a rather general CT aspect as far as fairness and professionality is concerned ("professionally and fairly"). This shows also that he is able to raise questions beyond the given information by referring to his knowledge about test design and characteristics of test quality.

Not unlike Alex' answer to other problems in the interview, his thinking related to this problem is marked by mathematical considerations: ST appears to be dominant. When reviewing the results of the CT analysis *under a combined ST/CT perspective*, the CT skills visible in the answer of Alex appear to be originated in ST knowledge Alex uses in his answer. The statement in the first sentence might be an exception: fairness and professionality of the study could mean that the people who did the study were not dependent from a computer brand and had no personal preferences, which might be an expression of rather general CT. However, the statement could also mean that the sampling was done in a correct, e.g. a randomized way – which would be more close to specific ST than to non-statistics-specific general CT. In the last sentence, Alex clearly refers to knowledge relevant for ST ("check, if possible, the significance").

Drawing a conclusion from the point of view of evaluating claims, our analysis yields that Alex' strength in ST allows him to use multiple strategies of questioning claims, mostly rooted in conceptual knowledge relevant for statistics.

We will now turn to the answer by Dana, who is also an experienced secondary teacher from Israel, and who evaluated the given statements differently:

Dana: Well, each of these statements can be true. I'll start with the first one: people – and me included in them – tend to listen to our friends' recommendation. Even though we read all the recommendation, we tend to feel our friends' experience more real. Also I think we tend to slightly disbelieve general recommendations even if they come from highly recommended magazines or web pages.
This second statement, while it sounds more convincing, it has a few issues. The first one being the question – which was not answered in this paragraph – if this magazine in a reliable one. The second one is if there is some personal gain to the people doing the test from one laptop manufacturer or not. And the biggest gap in the information here is what kind of malfunctions, based on what kind of use were these laptops put to.
I think the third statement is the more accurate one. Assuming both laptops are generally good, malfunctions happen on any electrical appliance. As I stated the type of malfunctions was not stated in the article and so it might be due to some use that Mrs. Blum will not require.
I believe that since people tend to go by a hunch, or, should we say, good feeling, they will go with what their friends recommend and so the first statement is the one I agree most with.

From the point of view of the *ST analysis*, Dana starts her answer with the claim that "all three statements can be true". This appears as somewhat inconsistent, as the given statements partly contradict each other. Her comment on the first statement focuses on the source of the recommendation and compares it, but does not make direct use of statistical considerations nor mention sampling issues. This also appears to be the case for Dana's comments about the second statement: She refers to the environment in which the computer magazine study has been conducted, however she does not mention its statistical power or sampling issues. Dana refers to the third given statement as "the more accurate one". This shows her awareness of uncertainty and chance.

Dana concludes that she agrees most with the first statement. Even though she had called the third statement "more accurate" she describes that people would rather rely on their friends in her view, from which she concludes that she would also prefer the first statement "and so the first statement is the one I agree most with".

Considering the *results of the CT analysis*, Dana shows aspects of CT skills in her very structured answer: At first, she admits that "each of the statements can be true" which may mean

that seen under a corresponding perspective, all the statements may be supported – not all at the same time but each of them may be selected as a possible point of view. This shows that Dana can think of other thinkers and their views, she appears to aim at first taking a neutral point of view. This approach may be seen as an aspect of CT.

In Dana's evaluation of the first statements she argues with feeling "we tend to feel our friends' experience more real". The "we" she uses probably refers to all human beings – this interpretation is also consistent with the expression "people" used by her in the last sentence. She describes what humans in her view tend to believe. This is however not questioned, Dana does not present an opposed view, she appears to follow her assumption of the people's preferences with her own judgment (see also her last sentence), which would be a very substantial deficit in CT.

In contrast, Dana's evaluation of the second given statement shows high CT skills: She questions the reliability of the source ("if the magazine is a reliable one", "personal gain […] from the laptop manufacturer") and the terms used in the statement or the information given in the problem ("what kind of malfunctions", "what kind of use").

She comes back to the "type of malfunctions" also in one of the subsequent sentences. This sentence follows Dana's evaluation of the third statement: "As I stated the type of malfunctions was not stated in the article and so it might be due to some use that Mrs. Blum will not require." This part of Dana's answer shows awareness of facts beyond the given data. However, even if she uses the word "might", the sentence prepares her conclusion in the last sentence, which is rather non-critical towards the "people's" judgment and the friends' recommendations. Dana's CT skills are hence very selective in this case, which, seen as a whole might stem from a premature, non-expressed judgment she might have made prior to giving her answer.

We now focus on the *combined CT and ST analysis*: As evidence of ST skills is rather absent in Dana's answer, she might have concentrated to rather "general" CT aspects which she used however, unsystematically, when evaluating the given claims and her own assertions or claims. The CT aspects of reliability of data sources (applied to the computer magazine) might have somewhat "blocked" a more careful consideration of ST-relevant aspects of the problem, such as the sample size. We may thus conclude that a few non-ST-specific strategies of evaluating claims have been used by Dana, whereas she did hardly use any strategy of evaluating claims which could be seen as ST-specific. CT in the intersection domain with ST appears to be lacking.

DISCUSSION AND CONCLUSIONS

The results replicate not only that a combined CT and ST analysis can better explain the evidence (cf. Aizikovitsh-Udi, Kuntze & Clarke, 2012), but the interpretation also yields insight into how CT and ST may interact in thinking processes. In particular, we have evidence on cases in which general CT is applied to the problem context without considering any use of relevant mathematical or statistical models. In such cases, this general CT did not interact with ST, and in some cases it even appears to somewhat block ST. However, as such CT may still contribute to critically evaluate the problem context, such CT could be classified as "CT enriching ST".

In contrast, we also observed cases in which the use of ST when evaluating interpretations of data could be interpreted as an adequate use of a CT strategy. Alex, for instance made such use of CT, often very much bound to his ST. We may hence qualify such CT as "CT within ST", as it is sometimes necessary for adequate ST. There are even algorithm-like procedures which implement CT elements in ST when interpretations of data have to be evaluated (e.g. when checking sampling characteristics or statistical significance).

We thus can conclude that there are components of CT which can be expected to have a visible influence on ST – and hence on SL: When evaluating claims in the form of interpretations of data, such "CT within ST" is crucial and an inclusion of a CT perspective can help to better distinguish such thinking steps.

Moreover, negative effects are not unlikely to occur: a strong focus on unspecific ST may block ST-relevant strategies of evaluating claims if it is combined with a thinkers' inclination not to question own claims made in the process of working on the problem. Moreover, the data suggests that a sufficient ST-relevant knowledge base as well as skills to use this knowledge in CT processes is a precondition of reaching higher levels of SL.

For teaching and learning, the findings hence suggest that an explicit combined focus on ST and CT could support the development of the learners' SL and avoid that a dominance of either ST or CT blocks thinking strategies in the corresponding other domain.

REFERENCES

Aizikovitsh-Udi, E., Kuntze, S., & Clarke, D. (2012). Connection between statistical thinking and critical thinking - a case study. In D. Ben-Zvi & K. Makar (Eds.), *Teaching and learning of statistics* (pp.239-249). *Proceedings of Topic Study Group 12, 12th International Congress on Mathematical Education (ICME-12)*. July 8-15, 2012, Seoul, Korea

Bullock, M., & Ziegler, A. (1994). Scientific thinking. In F. Weinert & W. Schneider (Eds.), *The Munich longitudinal study on the genesis of individual competencies*. Munich: MPI.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*(5), 382-393.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: Heath.

Dunbahr, K., & Klahr, D. (1989). Developmental differences in scientific discovery strategies. In Klahr & Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, New Jersey: Erlbaum.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). NY: Freeman.

Ennis, R. R. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, *18*, 4–10.

Ennis, R. H. (2002). Goals for a critical thinking curriculum and its assessment. In A. L. Costa (Ed.), *Developing minds* (pp. 44-46). Alexandria, VA: Assoc. for Supervision and Curr. Devel..

Ennis, R. H., & Millman, J. (2005). *Cornell critical thinking test, level Z* (5th ed.). Seaside, CA: The Critical Thinking Company.

Kröpfl, B., Peschek, W., & Schneider, E. (2000). Stochastik in der Schule: Globale Ideen, lokale Bedeutungen, zentrale Tätigkeiten. *mathematica didactica*, 23(2), 25-57.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674 - 689.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* San Diego, California: Academic Press.

Kuntze, S., Lindmeier, A., & Reiss, K. (2008). "Using models and representations in statistical contexts" as a sub-competency of statistical literacy – Results from three empirical studies. *Proceedings of ICME 11*. http://tsg.icme11.org/document/get/474

Kuntze, S., Engel, J., Martignon, L., & Gundlach, M. (2010). Aspects of statistical literacy between competency measures and indicators for conceptual knowledge. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of ICOTS 8*. Voorburg, NL: ISI. www.stat.auckland.ac.nz/~iase/publications.php [Refereed paper].

Lipman, M. (1991). *Thinking in education*. New York: Cambridge University Press.

Mayring, P. (2000). Qualitative Inhaltsanalyse. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [Online Journal],1*(2). Retrieved on 23rd November 2012 from http://qualitative-research.net/fqs/fqs-d/2-00inhalt-d.htm

McPeck, J. (1981). *Critical thinking and education*. New York: St. Martin's Press.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *The second handbook of research on mathematics teaching and learning* (pp. 957-1010). Charlotte, NC: Information Age Publishing.

Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1-8.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3-46.

Wild, C., & Pfannkuch, M. (1999): Statistical thinking in empirical enquiry. *International Statistical Review*, *3*, 223-266.