



**HAL**  
open science

# Decoding and discussing part-whole relationships in probability area models: The role of meaning-related language

Monika Post, Susanne Prediger

► **To cite this version:**

Monika Post, Susanne Prediger. Decoding and discussing part-whole relationships in probability area models: The role of meaning-related language. Seventh ERME Topic Conference on Language in the Mathematics Classroom, Oct 2020, Montpellier, France. hal-02970604

**HAL Id: hal-02970604**

**<https://hal.science/hal-02970604>**

Submitted on 18 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Decoding and discussing part-whole relationships in probability area models: The role of meaning-related language

Monika Post and Susanne Prediger

TU Dortmund University, Germany

monika.post@math.tu-dortmund.de prediger@math.tu-dortmund.de

*Although the epistemic role of academic language for developing conceptual understanding has been shown for some mathematical concepts, more topic-specific research is required for specifying the academic language demands for students' meaning making processes in different mathematical topics. The design research study presented in this paper contributes to this research agenda by investigating tenth graders' learning pathways towards conceptual understanding of conditional probabilities in area models. The qualitative analysis of a design experiment shows that students' processes of decoding and discussing meanings of different probability statements require a systematic scaffold for explicating the underlying part-whole relationships: the area model is helpful, but not sufficient.*

*Keywords: topic-specific language demands, probability area model, connecting multiple representations, explicating relationships*

## **Background: Epistemic role of language for the part-whole relationship in the mathematical topic of conditional probabilities**

### **Epistemic role of language in learning to discuss structural challenges**

Since Vygotsky's early writings about thinking and speaking (1934/1962), the epistemic role of language as a thinking tool has been emphasized by many researchers (e.g., Pimm, 1987): Language is not only relevant in mathematics classrooms as a means for communication but also for the mental processes of knowledge construction. Whereas procedural knowledge can be developed also with restricted language resources, the epistemic role of a school academic language has been shown to be particularly crucial for constructing conceptual understanding, e.g. for functional relationships (Prediger & Zindel, 2017). This and other studies indicate that a major challenge seems to be the articulation of abstract relationships.

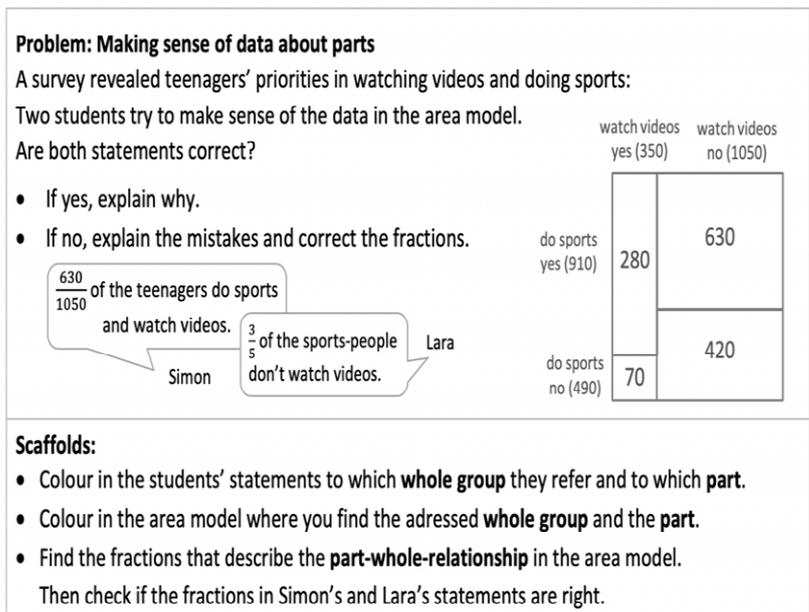
These general insights into the epistemic role of academic language for constructing conceptual understanding have immediate impacts for instructional designs: If we want to support language learners' development of conceptual understanding, we should identify those language demands which are crucial in this conceptual learning process and organize systematic learning opportunities so that all students can participate in discourse practices of explaining meanings (Prediger & Zindel, 2017). Although some aspects of this identification are generic for all mathematical topics, the detailed specification for a specific mathematical topic also requires empirical research (ibid.). Thus, the topic-specific identification of relevant meaning-related language demands has become a substantial step in the topic-specific design research agenda of the MuM research group in Dortmund. In this paper, we contribute to this research agenda for the mathematical concept of conditional probabilities and the underlying part-whole relationship in Grade 10/11.

**Topic in view: Conditional probability and the probability area model**

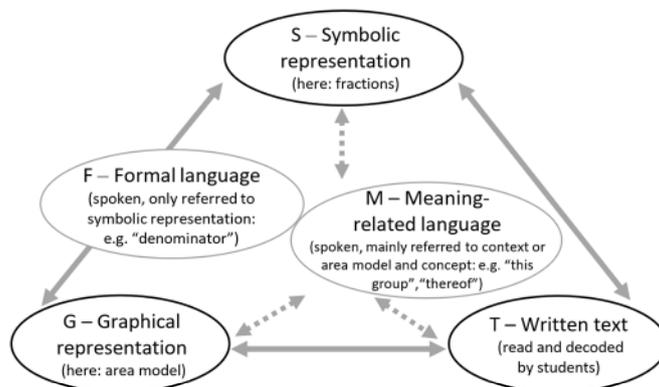
Conditional probability and Bayes’ rule have been shown to raise difficulties for many students. Already Shaughnessy’s (1992) research overview summarizes typical misconceptions and challenges, e.g., determining the conditioning event or the confusion between a conditional and its inverse or the challenge to distinguish conjunctive and conditional probabilities (as in the example in Figure 1). Within the succeeding decades, researchers in stochastics education developed and investigated approaches for fostering students’ reasoning, especially by making the structure of nested sets

more explicit for students (Sloman, Over, Slovak, & Stibel, 2003), e.g., in an area model as in Figure 1. Based on this research tradition, Böcherer-Linder and Eichler (2017) investigated which representations best support students’ access to complex part-whole relationships. By comparing students’ success rates of task completion for different representations, the authors showed that the area model and the tree diagram provide similar support for students in simple cases, but the area model is much more helpful for when horizontal sub set relations (i.e. Bayes’ rule) are addressed. Other researchers have investigated different complexities on the language side; for instance, Watson and Moritz (2002) found that students master statements about natural frequencies better than statements about probabilities and track this back to the complexity of the grammatical structure of the statements, particularly the phrase “out of” proved to be more accessible.

From these lines of research, we conclude that decoding the part-whole information can raise different complexities for different formulations and grammatical structures. Decoding the part-whole-information also seems to be challenging in written texts that contain no probability but only fraction statements such as in Figure 1. Whereas the existing literature suggests that decoding the nested structure of sets, the parts and the wholes, seem to be the crucial step in mastering conditional probabilities, this paper shows that the *part-whole relationship itself* is equally important for the process, not only the part and the whole and their nested structure. From the existing literature, we draw the need to visualize the parts and wholes, and will show that language is required in its epistemic role to focus the part-whole relationship besides the sets themselves. While existing studies have mainly focused on the comparison of students’ performance in different tasks or items and indicated what kind of task presentation could increase students’ access, we focus on developing students’ conceptual understanding and problem-solving strategies to master also complicated texts. In this design research perspective, the scaffolds printed in Figure 1 do not serve as the permanent support in each problem but as a strategy to be internalized for students’ independent access.



**Figure 1: Exemplary problem on decoding conditional probabilities (or fraction statements) in area models**



**Figure 2: Connecting multiple representations, with formal and meaning-related language**

This additional design ambition can build upon the previous research in stochastic education which resonates with the *general design principle of connecting multiple representations and language registers*. The principle has proven successful for developing language learners’ conceptual understanding (Prediger & Zindel, 2017) and is applied in our design. Figure 2 shows the representations in view for this article for the specific topic of conditional probability. We choose the area model with natural frequencies as the graphical representation that students are supposed to connect to the written text of the problem and the symbolic-numerical representation of fractions. Formal language means are those which refer to the symbolic representation. Meaning-related language means are all utterances that refer to the meaning as part-whole relationship (e.g., “group of sports-people”, “part”, “thereof”). The empirical section of this paper will show that the distinction between formal and meaning-related language is more crucial here than the registers because students require the meaning-related language as the epistemic tool for promoting students’ conceptual understanding for the distinction of conditional and conjunctive probabilities, prepared by fraction statements such as in Figure 1.

### **Research gap: Language demands in students’ learning pathways towards conceptual understanding of the part-whole relationship**

Although the area model with natural frequencies has already been identified as the most accessible graphical representation for simplifying students’ access to complex fraction statements in conditional probabilities, little is known on how to foster students’ ability to decode the structures independently, and on language demands that occur on this pathway. To explore the role of language as an epistemic tool in this process, the design research pursues the following research question: *Which academic language demands do students meet while developing conceptual understanding for conditional probability in the area model?* Once these language demands are identified, language support can be designed and provided throughout the conceptual learning trajectory.

### **Methodological framework: Data gathering and qualitative data analysis**

The methodological framework chosen for this project is design research because it combines two aims: designing a teaching-learning arrangement and developing an empirically grounded local theory of students’ learning pathways and the demands they meet (Gravemeijer & Cobb, 2006). For data gathering, four design experiment series were conducted in four classes in Grades 10 or 11 with 94 students between 15 and 18 years old. All sessions were video-recorded, a selection was

transcribed. The episode presented in the empirical part of this paper stems from Cycle 1 in which a class of 24 students worked on the teaching unit for 12 sessions of 60 to 75 minutes each. This episode from Cycle 1 was chosen to show best the epistemic role of meaning-related language.

In order to identify students' language demands while learning to decode written texts and while discussing part-whole relationships, the transcripts were qualitatively analyzed in three steps of deductive coding: In Step 1, each of students' utterance was coded when addressing the part, the whole or the part-whole relationship. Step 2 identified in which representations the components were addressed (with the letter codes S, F, M, G, T according to Figure 2). Correctly drawn connections are symbolized with continuous black lines, wrong connections with crossed lines, and not addressed connections are marked in grey. For distinguishing students' formal and meaning-related language, the interpreter analyzed how the students articulated the elements and their relation. All utterances referring only to the symbolic representation ("numerator", "the number above", articulation of numerator by rephrasing labels "350 are those who watch videos") were subsumed as formal language F. All utterances expressing references to the context of the data ("sports-people", "this group here") or the area model ("this rectangle") and at the same time referring to the underlying meaning-related concept/idea of fraction as part-of-the-whole ("thereof") were coded as meaning-related language M. Many utterances were double coded as F and M when referring to both. Interrater coherence was assured by consensus between two researchers. In Step 3, the codings of each sequence were graphically summarized (e.g., in Figure 3 and 4). The graphical summaries reveal what is made explicit or left unattended in a sequence. Hereby, we identify the function of the area model and the two languages and identify language obstacles in developing conceptual understanding.

### Empirical insights into decoding and discussion processes

The presented episode shows the whole class discussion of a Grade 10 classroom after the students' group work on the problem in Figure 1.

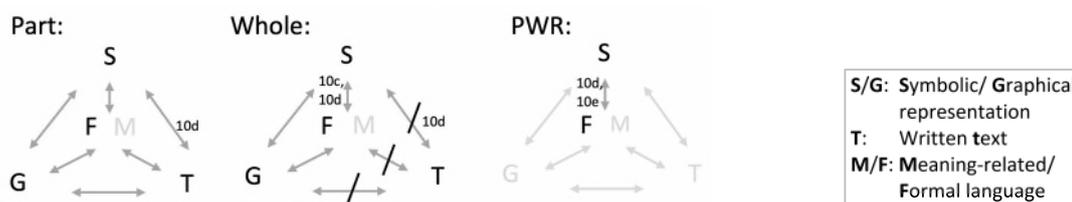
**Sequence 1: Celina's separate look on part and whole.** Sequence 1 starts when Celina presents her group's ideas about the wrong fraction statement of the fictitious textbook student Simon.

10 a	Celina	The first statement was, um, that $\frac{630}{1050}$ of the teenagers do sports and – at that – watch videos	
10b		And, starting with explaining the fraction, this statement is wrong, actually, because ... Um, first, um, the 630 you can see here [ <i>hints to the number 630 in the area model</i> ]. They do not watch videos and do sports [ <i>hints to the labels of the area</i> ]. And, sort of, that does not fit, first, because, um, sort of, because it is about, if teenagers watch videos [ <i>hints to Lisa's statement</i> ]	<b>Part:</b> S-G-F-T (numerator 630)
10c		And, um, to the second part of the fraction. These are, though, the teenagers who – sort of – do not watch videos, in total.	<b>Whole:</b> S-F-T (denominator)
10d		And the corrected fraction would be, um, I noted $\frac{280}{350}$ . Because, um, sort of, you should colour the important things in the text, and I have coloured "do sports" and "watch videos" [ <i>hints to both text fragments</i> ]. And, I wrote that 280 people watch videos and – at that – do sports [ <i>hints to the upper left rectangle 280 and their labels</i> ] And – sort of – this 350, these are all who watch videos then [ <i>hints to both left rectangles</i> ]	<b>PWR:</b> S-F <b>Part:</b> T-S-F-G (correctly identified) <b>Whole:</b> T-S-F-G (false-ly identified in T)

10e            That means, 350 are those who watch videos and the 280 those, those who – at that, even also - do sports [*hints the 280 in upper left rectangle*]. So I got  $\frac{280}{350}$ .

<b>Whole:</b> S-G
<b>Part:</b> S-G
<b>PWR:</b> S

Celina starts by explaining the fraction (focus on symbolic representation S) and by locating the numerator in the graphical representation (S-G-F-T for numerator in Turn #10a/b, see Figure 3). She correctly identifies that Simon connects the numerator to the wrong group in the text and the graphical model. From #10c on, she does not explicitly refer to the graphical model but only to its labels which also appear in the text (S-F-T for denominator in #10c). She corrects the fraction (addressing the part-whole relationship solely in the symbolic representation with the formal word fraction S-F in #10d), but does not verbalize the part-whole relationship (abbreviated PWR) beyond the word “at that” (“dazu” in German original). For explaining her choice of the fraction, she only focuses the part and the whole separately, but not the PWR. Her main concern is to express the nested structure of the conditions “watch videos” and “do sports and watch video”, but she neither expresses the part-whole relationship between the two, nor she realizes that the whole in view should be all teenagers. Thereby, she confuses the conjunctive with the conditional fraction statement.



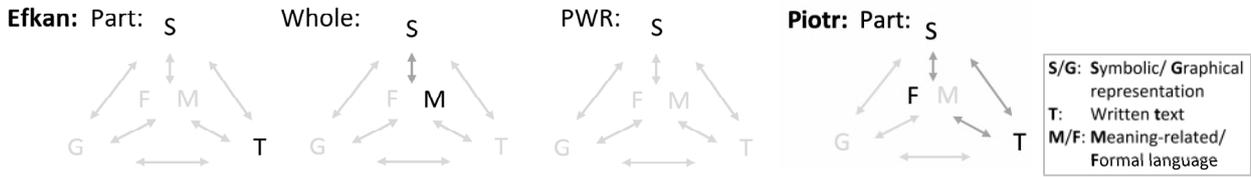
**Figure 3: Graphical summary of the analysis of Sequence 1: Celina’s separate focus on part and whole**

Figure 3 gives the graphical summary of her focus and expressions. The PWR is hardly addressed and this is characteristic of the sequences. Her own meaning-related language beyond rephrasing parts of the texts or labels is not activated, for none of the part, whole, or the PWR.

*Sequence 2: Meaning-related language focusing on part and whole.* Sequence 2 immediately follows Sequence 1 when Piotr and Efkán react to Celina’s statements.

12	Efkán	Well, I first had written $\frac{280}{1400}$ . Because, there are 1400 students, altogether, and 280 – as Celine has already said – do sports and watch videos. But, um, I still go hold on with my statement.	<b>PWR:</b> S <b>Whole:</b> S-M (“altogether”) <b>Part:</b> S-T
19	Piotr	I have had another idea, that is: 630 of 1050 of the teenagers do sports, but do not watch videos. Would do this there. There are 630 who, um, do sports but don’t watch videos. And there, we have “do watch videos”. I would only correct the statement, “but do NOT watch videos”.	<b>Part:</b> S-F-T

Piotr (in #19) seems to continue Celina’s pathway by starting with the fraction and only correcting the reference in the written text, but still explains a conditional fraction statement rather than a conjunctive one. In contrast, Efkán focuses the correction of the whole and succeeds in rephrasing “of the teenagers” in his own, meaning-related words (“students, altogether”, #12). He identifies the relevant groups, but still does not verbalize the part-whole relationship (Figure 4).

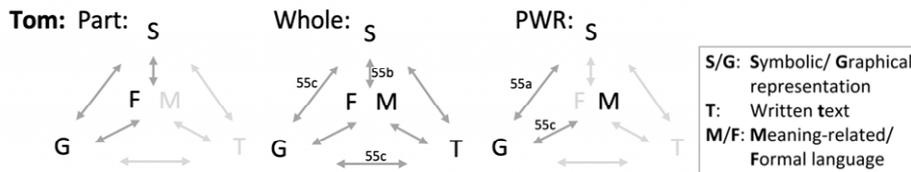


**Figure 4: Graphical summary of the analysis of Sequence 2: Efkan and Piotr focus differently**

*Sequence 3: Identification of whole by focusing on syntax and explicating the relation.* A little later, Tom presents his group work results for the second statement (of Lara).

- |     |   |  |
|-----|---|--|
| 55a | Tom Yes, well, $3/5$ is just this here, simplified. This would be $3/5$ and this $2/5$ [hints to the lower right rectangle]. And this would be the denominator, in this case [hints to the label “watch videos no”] and these both the numerators [hints to the two right rectangles].  | <b>PWR:</b> S-vaguely G<br><b>Part:</b> G-S-F<br><b>Whole:</b> G-S-F |
| 55b | And, um, that, well, you just have only this group here, considered [hints to the label “watch videos no”], so those who do watch no videos, but not those who watch videos. And sport, so all the persons [hints to the left rectangles], these are not considered in the denominator.   | <b>Whole:</b> G-M-S-F  |
| 55c | And, um, it says “ $3/5$ of the sports-people” [hints to the text], that means, you would need to consider them here [hints to two upper rectangles] and not these here [hints to two right rectangles]. That means, the denominator should actually be 910 of the sports-people and THEREOF, then 630, the numerator [hints to the upper right rectangle]. | <b>Whole:</b> T-G-S-F-M<br><b>Part:</b> S-F-G<br><b>PWR:</b> M-G     |

Tom starts with connecting the symbolic and graphical representation for the part and the whole. His connection to the graphical area model is realized only by deictic means, e.g., “this would be” in #55a. In contrast to Celina, Tom does not only verbalize the fraction, instead, he articulates the whole in own meaning-related words (in #55b) and connects them to the symbols. This allows him to connect formal and meaning-related language. With “thereof”, he also offers an explicit articulation of the part-whole relationship, supported by gestures on first the whole and then the part in a connection.



**Figure 5: Graphical summary of the analysis of Sequence 3: Tom’s full reconstruction of the whole**

### Discussion and outlook

The comparison of the three graphical summaries reveals typical profiles which we have also found in the analysis of many other episodes: Celina picks two conditions out of the text and interprets them as standing for numerator and denominator of the fraction. Thus she connects symbolic, graphical and textual representations for part and whole without searching for textual indicators of the part-whole relationship or for signifying the whole (part, whole: S-T-G-F, PWR: S-F). Hence, she adopts a direct translation strategy which often occurs in students’ comprehension processes and frequently results in misconceptions. Tom does not only pick out conditions, he also pays attention to the grammatical structure and the relationship between part and whole. Accordingly, he identifies the correct whole and justifies how to identity the whole in the text by the grammatical features of

genitive construction (translated to English by “of”). These mental processes are reflected by students’ use: Celina uses no meaning-related language besides the phrases given in the text and the labels in the area model (which are therefore not counted as own meaning-related language). This hinders her building an adequate model of the situation. Tom, in contrast, can articulate part (S-G-F), whole (S-G-FM-T) and the part-whole relationship (S-M-G) which gives him access to decoding and negotiating the adequate part-whole relationship.

Like Celina, Piotr also only concentrates on correcting the part without recognizing the wrongly decoded whole and without meaning-related verbalization of the part-whole relationship. In contrast, Efkan can recognize the wrongly decoded whole (S-M for whole; S-T for part) and expresses the whole using meaning-related language, but does not explicate the part-whole relationship. In his articulations, it becomes evident that the fact that the PWR has no direct counterpart in the graphical representation (being the relation between two rectangle areas) might be part of his challenge.

Especially the comparison between Tom and Celina shows: For distinguishing fraction statements with different structures and different referent wholes, it is crucial to articulate not only the part and the whole, but also the part-whole relationship in order to be able to express it, reason and negotiate about it. As long as the students do not find a language for expressing the part-whole relationship, the direct translation strategy for the separate conditions has a certain instrumental rationality. However, the empirical insights into these and also other episodes repeatedly show the high relevance of language, here in its epistemic function of allowing students to internally construct adequate models of the situations and models for distinguishing conditional and conjunctive fraction statements.

The observation that some students (as Piotr in #19) cannot follow other students’ reasoning (such as Efkan’s in #12) strengthens the need to support students in developing their meaning-related language for identifying parts, wholes, and particularly part-whole relationships. The area model alone provides a helpful scaffold for the part and the whole, but not really for the part-whole relationship. Hence, language – in these cases – appears as crucial not only for the textual representation of the situations but mainly for students’ thinking processes during decoding and discussing. We therefore deem it necessary to compile an inventory of all phrases that students used to express the rarely articulated part-whole relationship. Including observations from other episodes of the data material, Figure 6 summarizes options of increasing density for articulating the part-whole relationships in meaning-related ways. These kinds of phrases were identified as those language means for which most students require focused language-learning opportunities.

Articulation of PWR with increasing density	Two-step phrase, with “thereof”	Two-step phrase with “out of them”	Connective “(out) of”	Genitive construction (not translatable)	Condensed term “share”
Examples	There are 91 sports-people altogether and thereof 63 don’t watch videos.	There are 54 female persons and 45 out of them do sports.	Out of the girls 45 do sports. $\frac{52}{75}$ of the girls do sports.	(German original: 52 der 75 Mädchen ...)	Share of sports-men among the boys is $\frac{20}{90}$

**Figure 6: Meaning-related phrases for articulating the part-whole relationship in the probability area model: Inventories from several design experiments (literally translated from German)**

Of course, the empirical findings of the case study must be interpreted with respect to their methodological limitations, most importantly (a) being bound to the specific teaching-learning arrangement in view, (b) the small sample size, and (c) the missing language support in the currently investigated Design Experiment Cycle 1. Future research should extend these boundaries.

However, the empirical findings have substantially informed the re-design of the teaching-learning arrangement which now includes the structural scaffolds as already printed in Figure 1 (which explicitly ask for the part-whole relationship and not only for part and whole separately) and a lexical work on meaning-related phrases for expressing it in the different contexts has been included by providing meaning-related language frames to verbalize the graphically given part-whole relationships. Once again, the study could replicate the high relevance of the epistemic role of language, especially for expressing and thinking about abstract relationships. Whereas concrete objects and sets can be visualized easily, understanding the abstract relationships requires their verbalization, and this phenomenon has now been reconstructed for the part-whole relationship in probability area models as well. Rather than simplifying all texts until the part, the whole and the PWR are easily decodable (as many of the existing studies in the field of stochastics education might suggest or be misinterpreted to suggest), the ultimate aim of language-responsive classrooms must be to equip all students with the language means to master also difficult demands. Before reaching this aim, further research is required.

### **Acknowledgement**

This paper was developed in the MuM-research group, financially supported by the German Ministry for Education and Research BMBF and Deutsche Telekom Foundation.

### **References**

- Böcherer-Linder, K., & Eichler, A. (2017). The impact of visualizing nested sets. An empirical study on tree diagrams and unit squares. *Frontiers in Psychology*, 7:2026, 1–11.
- Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 17–51). London, UK: Routledge.
- Pimm, D. (1987). *Speaking mathematically: Communications in mathematics classrooms*. London, UK: Routledge.
- Prediger, S., & Zindel, C. (2017). School academic language demands for understanding functional relationships. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(7b), 4157–4188.
- Shaughnessy, J. M. (1992). Research on probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York, NY: Macmillan.
- Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, 91(2), 296–309.
- Vygotsky, L. S. (1934/1962). *Thought and language* (E. Hanfmann & G. Vakar, Eds. and Trans.). Cambridge, MA: MIT Press.
- Watson, J. M., & Moritz, J. B. (2002). School students' reasoning about conjunction and conditional events. *International Journal of Mathematical Education in Science and Technology*, 33(1), 59–84.