2021

PROMPTING SELF-EXPLANATIONS DURING THE LEARNING OF PROBABILITY: CONTENT-SPECIFIC VERSUS GENERIC VERSUS GENERIC WITH A FORM OF GUIDANCE

Genghu Shi

PROMPTING SELF-EXPLANATIONS DURING THE LEARNING OF PROBABILITY: CONTENT-SPECIFIC VERSUS GENERIC VERSUS GENERIC WITH A FORM OF GUIDANCE

by

Genghu Shi

A Dissertation

Submitted in Partial Fulfillment of the

Requirement for the Degree of

Doctor of Philosophy

Major: Psychology

University of Memphis

August 2021

Acknowledgments

Firstly, I would like to express my deep appreciation and thanks to my mentors Dr. Xiangen Hu and Dr. Arthur C. Graesser for their continuous and tremendous support of my Ph.D. studies and related research. Their mentorship, guidance, and expertise will have life-long benefits for my future career and personal life. I would also like to thank them for their help and great suggestions when I hunted for an academic position in universities.

Meanwhile, I would like to thank my graduate committee of Dr. Philip Pavlik Jr., Dr. Jia Wei Zhang, and Dr. Su Chen not only for their time to help me defend my dissertation proposal and dissertation itself but also for their insightful suggestions on my dissertation research. Here, I would express my special thanks to Dr. Su Chen who mentored my graduate study in applied statistics. At last, I received my master's degree in statistics which lend me more credits when I hunted for a job.

I sincerely thank my fellow graduate students who always gave me support when I need help both in my life and in the research. They are Ying Fang, Keith Shubeck, Lijia Wang, Liang Zhang, and Meng Cao. This list should also include the lovely lady Jody Cockroft who is not a graduate student, but her help and warm smile will never be forgotten. I would definitely like to thank everyone appeared in my life during this long but short journal in my Ph.D. study in United States.

Finally, I would like to thank my wife Hong Shi, my parents and parents-in-law for their support in my life.

Abstract

Shi, Genghu. Ph.D. The University of Memphis. August 2021. Prompting Self-explanations during the Learning of Probability: Content-Specific, Generic versus Generic with a Form of Guidance. Major Professor: Xiangen Hu, Ph.D., and Arthur C. Graesser, Ph.D.

Learners often cannot apply (transfer) the knowledge they learned from instructional settings into a new context. Therefore, their knowledge is likely “inert.” Research shows that learners must be actively involved in learning construction activities to enable knowledge transfer to occur. Self-explanation is one such constructive cognitive activity that involves explaining learning materials (expository texts, worked examples) to oneself with attempts to make sense of new information. It has been shown to support deep comprehension and knowledge transfer. However, self-explanations usually cannot be spontaneously generated by learners, but need to be elicited by prompts. The prompts can range from generic type (e.g., “Explain this!”) to content-specific type (e.g., filling in the blank of an incomplete sentence or selecting an explanation from multiple choices.) based on the amount of guidance they provide. This dissertation investigated the effectiveness of three types of self-explanation prompts (content-specific prompts, generic prompts, and generic prompts with a form of guidance) being applied to learners with different levels of aptitudes (prior knowledge and learning ability) when they learn probability. The self-explanation session was implemented in AutoTutor. The learners were prompted to self-explain correct and incorrect solutions to procedural probability questions. Four research questions were investigated in the study. First, are all three types of prompts effective in improving learning? Second, are the generic prompts with a form of guidance more effective than content-specific and generic prompts? Do they elicit more high-quality self-explanations in general? Third, are there interaction effects between learners’ aptitudes and different types of prompts? And lastly, do high-quality self-explanations facilitate far transfer of knowledge? The results suggested that only generic prompts with or without guidance were effective in improving learning. Moreover, they were more effective but did not elicit more high-quality self-explanations than content-specific prompts. There were no interaction effects between learners' aptitudes and different types of prompts, which means that learners' aptitudes do not vary the effects of different types of prompts on learning. High-quality self-explanations predicted far transfer of learning, as was expected. The dissertation discusses the results, the limitations of the study, and future directions on self-explanation research.

Table of Contents

List of Tables

List of Figures

Prompting Self-explanations during the Learning of Probability: Content-Specific versus Generic versus Generic with a Form of Guidance

Introduction

Learners[1] often cannot use the knowledge they learned from instructional settings, such as schools, universities, and vocational institutes, in a new context (Kurtz & Honke, 2020; Renkl et al., 1996). The lack of knowledge transfer makes learners' knowledge stay "inert" (Bereiter & Scardamalia, 1985; Whitehead, 1955). For Learners to apply the inert knowledge in the real world, they must be actively involved in the knowledge construction activities (Bransford et al., 2000; Cote, 1994).

Self-explaining is one constructive cognitive activity that involves explaining learning materials (expository texts, worked examples) to oneself with attempts to make sense of new information (Chi, 2000; Rittle-Johnson & Loehr, 2017). So, what are the characteristics of self-explaining? According to Mitsea and Drigas (2019), it is a metacognitive learning strategy because it involves activities that externalize learners' thinking processes and the contents of the tasks. With self-explanation, learners can justify a decision or belief, explain a concept, make a prediction, or make a metacognitive judgment about their understanding, reasoning, and explanations (Bisra et al., 2018; McNamara, 2004). By systematic detailed analysis, Chi (2000) revealed its underlying mechanisms that, during self-explaining, learners monitor their understanding, generate inferences, and revise their incomplete or incorrect mental models of the

[1] This dissertation refers to "learner" or "learners" hereafter given that the targeting learners are not only school learners.

learning materials. The details of these mechanisms will be introduced in the *Literature Review* section.

According to Chi (2000), when learners detect gaps or missing information in learning materials, they generate inferences about causal connections among objects and events by integrating current information with prior knowledge. The causal connections make learning materials more comprehensible (McNamara et al., 2007). The inferences generated by learners are considered new knowledge they will construct in their knowledge structure or mental model (Chi, 2000). This is known as the inference-generating mechanism of self-explanation. However, Chi (2000) observed some interesting phenomena that the incomplete learning material view fails to explain. For example, some self-explanations are fragmented, and sometimes even incorrect, but they do not seem to harm learning. Obviously, these fragmented or incorrect explanations cannot be viewed as incorrect inferences. They must serve other purposes that the inference-generating mechanism cannot explain. In addition, researchers found that self-explanations seemed to be clustered at some key locations of the learning material, but these locations were not the sites where crucial information was missing. The revision view can supplement the self-explanation mechanism with the perspective that learners' gaps in their mental models of the learning materials exist. In this perspective, the fragmented and incorrect self-explanations may be caused by the learners’ flawed or imperfect mental models externalized during self-explaining. The key locations where the self-explanations cluster may be the crucial missing information in their pre-existing mental models. Thus, self-explaining is conceived as a process of self-revising learners’ existing incomplete and incorrect knowledge structure or mental models of the learning materials. However, both the gaps in the learning materials and learners' mental models need to be actively detected or monitored by learners during self-explanation.

Self-explanation is a powerful learning technique and is generally effective across various topics, such as mathematics (Aleven & Koedinger, 2002; Berthold et al., 2009; Rittle-Johnson et al., 2017), physics (Chi et al., 1989), biology (Chi et al., 1994), and law (Aleven et al., 2006). It also helps learners in different age groups from pre-school children to college undergraduates (Bisra et al., 2018; McNamara, 2004; Rittle-Johnson et al., 2017). Self-explaining supports comprehension and far transfer of knowledge (Chi et al., 1989; Rittle-Johnson & Loehr, 2017). In general, it activates the inert knowledge and enables the newly obtained knowledge to be applied in new contexts. In the seminal work of self-explanation, Chi and colleagues (1989) found that learners who had greater success in applying newly learned knowledge to solve problems tended to generate self-explanations spontaneously. Also, learners who spontaneously generated self-explanations provided better justifications for each step they took while solving the problem. In contrast, learners who performed poorly rarely explained the expository texts they read or their problem-solving steps. When they did, they could not generate causal connections among the principles and concepts.

Unfortunately, Chi's research also showed that very few learners spontaneously self-explain during learning, so they somehow need to be prompted to do so. Chi and colleagues (1994) compared the learning gains of learners who were prompted to self-explain a biology text passage with those who read the passage twice. They equated the time the two groups spent on the text and still found that the learners who were prompted to self-explain had greater learning gains than the control group. Further, among the prompted learners, those who generated a greater number of self-explanations, regardless of their accuracy, had a deeper understanding.

Two recent meta-analyses, conducted by Rittle-Johnson and Leohr (2017) and Bisra et al. (2018), revealed that prompted self-explanation is a highly effective and self-directed intervention for improving the learning of both conceptual (declarative) and procedural knowledge. They also suggested that self-explanation should be grasped by learners for self-regulated learning and scaffolded by learning environments. Together, these results conclude learners should be prompted to self-explain when they learn new knowledge.

While generally effective, prompted self-explanation has sometimes been observed to have negative or no effects on learning (e.g., Broers & Imbos, 2005; DeCaro & Rittle-Johnson, 2012; Kuhn & Katz, 2009; Mwangi & Sweller, 1998). The mixed results signify that the constraints on the effectiveness of prompted self-explanation exist (Rittle-Johnson & Loehr, 2017). For example, Kuhn and Katz (2009) suggested that, under some conditions, prompting to explain one's solutions or ideas may reduce the effectiveness of self-explanation. Learners may reduce their attention to new information when they repeatedly explain their preexisting mental model. Berthold et al. (2011) found that prompts that focused on key concepts increased conceptual comprehension, but reduced transfer. They argued that the conceptually oriented prompts draw learners' limited attention to the key concepts at the expense of neglecting the procedural knowledge. Williams et al. (2013) directly asserted that erroneous overgeneralizations caused by prompted self-explanation could be hazardous to learning. These constraints on the effectiveness of prompted self-explanation motivate the need to conduct more research to identify the conditions that prevent the negative effects of self-explanation and maximize and extend its benefits in practice.

Advances in learning technologies have progressed to incorporate computational models in artificial intelligence, learning sciences, cognitive sciences, and computational linguistics, including the development of Intelligent Tutoring Systems (ITSs, Graesser et al., 2018). ITSs are computer-based systems designed to provide individualized instruction to learners by modeling and adapting to individual learners' prior knowledge, behaviors, skills, affect, and mental states. After decades of efforts by researchers in this field, ITSs have been improved to a point where they are more effective than conventional instructions and significantly improve learning outcomes for learners (Kulik & Fletcher, 2016; Steenbergen-Hu & Cooper, 2014; VanLehn, 2011). ITSs have been widely deployed to millions of learners since their inception. Some have had significant impacts on education. For example, ALEKS (Assessment and LEarning in Knowledge Spaces; Canfield, 2001) has been successfully used in after-school programs to reduce the math gaps between white and black learners (Hu et al., 2013; Huang et al., 2016). The Cognitive Tutors for algebra have been implemented in thousands of middle and high schools and have yielded improvements in learning gains and speed of learning (Koedinger et al., 1997; Koedinger & Corbett, 2006; Ritter et al., 2007). iSTART (interactive strategy training for active reading and thinking) is a web-based ITS that provides adolescents to college-age learners with high-level reading strategy training to improve their comprehension of science texts (McNamara et al., 2004; McNamara et al., 2007).

Despite this progress, ITSs still have room to improve their effectiveness and efficiency. For example, none of the current ITSs can attain Bloom's 2-sigma challenge that ITSs and one-on-one human tutors should ideally be two standard deviations more effective than conventional classroom environments (Bloom, 1984; VanLehn, 2011). There is a long way for ITS researchers to go to narrow this gap. Meanwhile, ITSs provide a testbed for researchers to create more effective, intelligent, and individualized learning environments and a perfect experimental platform to explore the untouched or insufficiently investigated research questions in learning sciences.

This study explored the interaction effects of different types of self-explanation prompts and the aptitudes of learners who learned about probability. Here, the aptitudes refer to the learning ability and prior knowledge. The learning ability was defined as the amount of knowledge a learner retains and comprehends after learning a topic within a particular window of time. Prior knowledge was defined as the amount of declarative knowledge a learner retains after learning a topic within a particular window of time. The learners learned through a conversation based ITS called AutoTutor (Graesser, 2016). There are several reasons for this study. First, self-explanation has already been successfully implemented in ITSs (e.g., Aleven & Koedinger, 2002; Conati & VanLehn, 2000; McNamara et al., 2004; McNamara et al., 2007). This improved learners' learning even with simple prompts. These prompts included requests to select the name of a problem-solving principle from a menu to justify the problem-solving steps, or simply asking learners to fill in the blank of a partial definition of a problem-solving principle (Aleven & Koedinger, 2002; Conati & VanLehn, 2000). It is hard to say that these types of prompting were really eliciting self-explanation because simply recollecting the names or the content of the principles did not guarantee that learners made causal connections between a step of problem-solving and a specific principle. The learners might not be able to apply these principles to solve a problem. Later, Aleven et al. (2004) used a tutorial dialogue system to support learners expressing self-explanations in natural language dialogues. They did not find that the natural language explanations lead to better learning outcomes compared to simply selecting a principle's name from menus. In the discussion, they ascribed the undifferentiated results to the sampling artifact that their learners were all high-ability learners. However, they did find that learners learned better to state explanations with the support of the tutorial dialogue system compared to the learners who simply selected principle names from menus. It is important to point out that the self-explanation prompts used in these studies were all content-related or content-specific. These studies focused on eliciting the explanations of specific principles that were key to solve a problem. They did not use generic content-independent prompts (e.g., "Explain this!") or compare these two types of prompts. Currently, it is unclear how these two types of prompts impact learners with different learning abilities.

Second, very few studies, not only in the field of ITSs, compared these two types of self-explanation prompts (Aleven et al., 2006). Content-specific prompts may be helpful for some learners to realize that they have gaps in their understanding and get hints to fill these gaps (VanLehn et al., 1992). But such prompts may not benefit the learners who already understood these contents, or even worse, they may prevent learners from generating a series of inferences because they direct learners' attention to specific content (Aleven et al., 2006). On the other hand, Chi (2000) has claimed that generic prompts (e.g., "Explain this sentence to yourself") should be more effective presumably because they enable learners to tailor their self-explanations for revising their own incomplete or incorrect knowledge structure or mental model. Generic prompts increase the opportunity for learners to detect gaps in their understanding, discover deficiencies in their mental models of the learning contents, or generate useful inferences (Chi, 2000; VanLehn et al., 1992). Given that generic prompts, by definition, are not related to a specific domain, they can be implemented in different domains and systems with little to no editing (Kramarski et al., 2013).

No significant difference was observed between generic and specific prompts in the meta-analysis conducted by Bisra et al. (2018), but the study did not directly compare the two types of prompts. Chou and Liang (2009) reported in their study that learners performed better in near transfer problem-solving when they were prompted by both content-specific and generic questions. But there was no significant difference between the two types of prompts. Aleven et al. (2006) reported that, in ill-defined domains (e.g., legal reasoning), less able learners benefited more from content-specific prompts and more able learners learned better with generic prompts. As far as I know, there is no study specifically investigating what kinds of prompts (generic versus specific) are more effective in promoting learning gains of learners with different levels of ability (high versus low) in well-structured domains (e.g., probability). In the current study, learners with different ability levels will be asked to explain why the solutions to a problem are correct or incorrect with different types of prompts in the current study.

Self-explaining is typically more effective for learners when they receive a form of assistance alongside prompts (Berthold et al., 2009; Margulieux & Catrambone, 2019). The amount of guidance a prompt provides varies. Prompts can be completely open-ended questions that provide no guidance, such as “Can you explain that?” (Hausmann & Chi, 2002). They can also be focused questions that provide some guidance, such as “Explain how examples 1 and 2 are similar” (De Koning et al., 2011). The guidance provided in the focus questions directs learners’ attention to specific learning content. When providing a lot of guidance, prompts do not elicit as many constructive self-explanations, such as selecting explanations from a menu (Aleven & Koedinger, 2002) or filling blanks in partial definitions (Conati & VanLehn, 2000). Given these results, there is an open question on the effects of generic prompts with some guidance that is not directly related to or refers to the current learning content but related to the domain. For example, before being prompted to self-explain why a solution to a probability problem is right or wrong, learners are provided some guidance about the common errors that they may commit during problem-solving. The generic prompts with such guidance composes the third condition of the current study.

Third, probability is one of the most important subdivisions in mathematics. "*Probability is a way of thinking.*" (Falk & Konold, 1992). It not only helps people develop critical thinking in daily life (Aizikovitsh & Amit, 2008), but also is foundational in other disciplines such as statistics, psychology, physics, biology, medicine, business, and politics (Ang & Shahrill, 2014; O'Connell, 1999). However, a large proportion of learners enter and leave courses with misconceptions about probability, which prevents them from transferring or applying the probability concepts and principles into new contexts (Khazanov & Prado, 2009; Khazanov & Prado, 2010). Self-explanation may help learners overcome the misconceptions with its facility to support comprehension and far transfer (Chi et al., 1989; Rittle-Johnson & Loehr, 2017). Additionally, according to the National Council of Teachers of Mathematics (NCTM, 2000), mathematical communication should be an essential ability equipped by twenty-first-century learners learning mathematics. Mathematical communication here refers to sharing ideas and clarifying understanding. Overt self-explaining, such as typing or speaking out loud, may play a role in helping learners develop such an ability.

Fourth, AutoTutor is a conversation-based intelligent tutoring system that supports a mixture of vicarious learning and interactive tutoring (Graesser, 2016; Nye, Graesser, et al., 2014). In vicarious learning, human learners learn by observing peer agents asking deep questions and a tutor agent promptly answering each of these questions (Gholson & Craig, 2006). In interactive tutoring, human learners answer a main question by interacting with a tutor agent with multiple turns in natural language. AutoTutor provides an ideal experimental platform for learning science research. AutoTutor can be used to display learning materials (e.g., Google slides, images, or text), while conversational computer agents can be used to provide experimental instructions and guide learners to go through the experiments.

Testing the effects of different types of self-explanation prompts on learners of different levels of learning abilities make it possible for ITSs to provide learners with an individualized learning experience when they self-explain the learning content. Some ITSs have supported the positive effects of self-explanation on learning and transfer but have not contrasted the different prompts investigated in this dissertation. It is worthwhile to investigate how effective the three different kinds of prompts are for different categories of learners.

Literature Review

Definitions

Self-explanation

Self-explanation is a constructive learning strategy to make sense of new information by explaining to oneself (Chi, 2000). It helps learners construct new knowledge by elaborating on the learning materials, relating them to prior knowledge, making inferences, and making connections among given information (Chiu & Chi, 2014). In this sense, the activity of self-explaining is generated by and also directed to oneself. Self-explanation also refers to "a unit of utterances produced by self-explaining" (Chi, 2000). Self-explanation involves explaining information to oneself while learning, such as explaining texts while reading (Chi, 2000; McNamara, 2004), justifying the worked example solutions (Rittle-Johnson et al., 2017), and building connections across multiple representations of a piece of knowledge (Berthold et al., 2009), etc. Self-explanation has been included in the *25 Learning Principles to Guide Pedagogy and the Design of Learning Environments* from the Association for Psychological Science

(Graesser, 2008) and in the *7 Recommended Learning Strategies* in the 2007 Practice Guide of Institute for Educational Science (IES) (Pashler et al., 2007).

Types of Self-explanations

Chi (2000) and McNamara (2004) identified five types of self-explanation: elaborations, paraphrases, inferences, self-monitoring statements, and nonsensical statements. Elaboration is a strategy to connect different ideas or concepts, whether the connections are meaningful or not. It usually serves the purpose of memorizing the learning content. Paraphrase refers to repeating or expressing the content of a text, or in one's own words, without generating further information or new knowledge. An inference or a self-explanation inference (SEI) is a piece of new knowledge generated by integrating information across learning contents and prior knowledge. Other types of inferences, such as bridging, paraphrasing, logical inferences, and schema-based inferences, do not generate new subject-matter knowledge. These are not considered as self-explanation inferences (Chi, 2000). Self-monitoring statements are used to indicate learners' understanding or uncertainty of the learning content, e.g., "It is easy to understand", or "I don't know". Inferences and self-monitoring statements are high-quality self-explanations (Wylie & Chi, 2014), whereas others are noted as low-quality self-explanations. The number of high-quality self-explanations generated by learners is positively related to their learning gains (Chi et al., 1989).

The self-explanation effect

In the seminal work on self-explanation, Chi et al. (1989) found that learners who spontaneously generated more self-explanations (high explainer) when studying worked examples in physics learned better than those who generated fewer self-explanations. Regardless of the accuracy of self-explanations, just the act of self-explaining improved learning gains. This is known as *the self-explanation effect,* which has been replicated across domains such as biology (Chi et al., 1994), computer science (Pirolli & Recker, 1994; Recker & Pirolli, 1995), history (Wolfe & Goldman, 2005), legal reasoning (Aleven et al., 2006), and mathematics (Renkl, 1997; Rittle-Johnson et al., 2017). Meanwhile, research over the last three decades has demonstrated that self-explanation benefitted learners across a wide range of ages, from preschool children to college undergraduates (e.g., Bisra et al., 2018; Chi et al., 1994; Rittle-Johnson et al., 2017). These learners were instructed using different learning materials, such as worked examples, texts, texts with multiple representations, and even learning resources provided by ITSs (Aleven & Koedinger, 2002; Berthold et al., 2009).

As a constructive learning activity, self-explanation improves conceptual knowledge (e.g., Aleven & Koedinger, 2002; Berthold et al., 2009; McEldoon et al., 2013), procedural knowledge (e.g., Atkinson et al., 2003; McEldoon et al., 2013; Rittle‐Johnson, 2006), and knowledge transfer (see Catrambone & Yuasa, 2006; Chi et al., 1989; Renkl et al., 1998). Conceptual knowledge includes facts, concepts, and principles in a domain (De Jong & Ferguson-Hessler, 1996). Self-explanation improves conceptual knowledge by repairing and enriching existing knowledge structure to make it more accurate and better structured (Chi, 2009; Chi et al., 1989; Fonseca & Chi, 2011). Through this process, learners gain a deeper understanding of the principles and the relationships between units of knowledge (Rittle-Johnson et al., 2001). Procedural knowledge is often defined as knowledge of procedures (Rittle-Johnson et al., 2001; Star, 2005, 2007), a series of steps, or actions enacted to accomplish a goal. By self-explaining the steps of problem-solving, learners may gain insight into the rationale for a procedure, which in turn may broaden the range of problems that apply the procedure and ultimately promote procedural transfer (Rittle - Johnson, 2006).

A rich body of studies has supported the benefits of self-explanation to conceptual knowledge, procedural knowledge, and knowledge transfer (e.g., Aleven & Koedinger, 2002; Atkinson et al., 2003; Berthold et al., 2009; Bisra et al., 2018; Chi et al., 1989; McEldoon et al., 2013). For example, the study conducted by McEldoon et al. (2013) showed that self-explanation promoted both conceptual and procedural knowledge of mathematics, particularly knowledge of equation structures and transfer of the equation structures. Renkl et al. (1998) found that self-explanation fosters both near and far transfer of calculation problem-solving skills. Near transfer refers to solving the types of problems they practiced; far transfer refers to solving related but not isomorphic problems (Haskell, 2001). Catrambone and Yuasa (2006) demonstrated that prompting self-explanations yielded greater success at locating the relevant information needed to perform transfer tasks when utilizing computerized databases. Hilbert et al. (2008) further suggest that adding self-explanation to worked examples improves learners' conceptual knowledge of Geometry. Rittle - Johnson (2006) found that self-explanation helped children learn and remember a correct procedure and promoted procedural transfer by focusing on explaining the procedures but did not lead to greater improvements in conceptual knowledge. These findings suggest that self-explanation can improve learners' understanding of the underlying concepts inherent in the problems and their ability to carry out the steps and transfer them to new problems. However, further research is needed to clarify why self-explanation sometimes fails to promote conceptual knowledge when it should.

Generally speaking, generating self-explanations per se is useful in improving learning and transfer. However, researchers recommended that learners should actively engage in self-explaining to make it play its full role as a constructive activity (Bransford et al., 2000; Chi, 2009).

Two Mechanisms of Self-explanation

We cannot explain the mechanism or the internal process of self-explanation effect, and the individual differences during this process, by simply stating that self-explaining improves learning because it is a constructive activity. As the definition indicates, self-explaining is conceived as a constructive activity of generating inferences (Chi, 2000). Such inferences are generated to fill the omissions in the learning materials in which the texts are usually assumed to be incomplete no matter how well written. The “omissions” in texts are corresponding to the “gaps” in one’s mental model. In other words, self-explaining is a process of generating inferences to fill the missing information in the learning materials, which, in turn, fills the “gaps” in one’s understanding of the materials. This process results in a mental model that is isomorphic to the text model in the learning materials. The inference-generating view was first supported by the work of Chi et. al. (1994). However, this view failed to explain some findings in their research, such as why some learners generated more inferences than others and why learners did not uniformly explain sentences across the text since the “omissions” were distributed in some fixed positions of the text. Such unexpected findings suggest that self-explaining seems to serve another purpose, a purpose tailored to one’s own needs. By further analyzing the self-explanation inferences, Chi (2000) proposed another mechanism that assumes self-explaining is a process of repairing one’s flawed mental model. Presented below, the details of the two perspectives of self-explaining are presented.

Inference-generating View

The assumption behind this view is that self-explanation is a process of generating inferences beyond the information contained in instructional material is that such material is incomplete in some way, no matter whether it is poorly or well-composed (Chi, 2000). For example, a poorly written text or a worked example may be structurally and explanatorily incoherent. Structural incoherence happens when anaphoric references and/or connective ties are missing between sentences, thus destroying the structural coherence. Kintsch and Vipond (2014) found that structural coherence facilitates comprehension in general for all learners. Explanatory incoherence happens when some pieces of crucial background information are left unstated (Kintsch & Van Dijk, 1978). Providing such crucial information by generating inferences will no doubt benefit learners (McNamara et al., 1996). On the other hand, even well-composed material, e.g., textbooks, can fail to convey all information about a topic. Learners must generate inferences to fill the gaps in their understanding of the topic in the external form of filling the omissions in instructional materials. So, the more inferences a learner generates, the more enriched his/her mental representation is, and the better he/she can learn.

As we know, the main goal of self-explanation is not just to make poorly composed learning material coherent, but to generate inferences that infer new knowledge, improves learning new domain knowledge. Chi (2000) postulated several inference-generating mechanisms that explain how self-explanation helps with learning a new domain without prior knowledge. First, one can generate inferences by integrating information from different parts across the instructional material. Second, one can generate inferences by using analogy or comparison to integrate information presented in the instructional material with commonsense or domain-related knowledge. Third, one can use the meanings of words in the instructional material to imply what may also be true. Of course, inferences can also be generated by combining any of the three inferencing mechanisms.

. However, the inference-generating view of self-explanation failed to explain some unexpected findings in the study of self-explaining biology text by Chi et. al. (1994). Their study found that self-explanation inferences generated by learners did not always make sense. They were often fragmented, or even incorrect. By further coding the inferences, it was revealed that, although 25% of the self-explanations were erroneous, the learners nevertheless learned from generating them. The fact that erroneous self-explanations were not detrimental to learning suggested that they may serve another purpose. Two other findings from this study could not be explained by the inference-generating view either. First, learners' self-explanations were not uniformly clustered at the same sites of the text passage. The information omissions were scattered in the text passage, so the assumption would be that each learner's self-explanations should mostly happen at the locations of the information omissions. However, learners not only did not have a consensus in the loci of their self-explanations, but they also did not often generate self-explanations at the sites where crucial information was missing. The second unexplained finding was that learners' self-explanations at the same location were not always semantically equivalent. When combined, findings suggest that self-explanation seems to serve another purpose tailored to one's own needs. In this alternative perspective, it is presumed that learners come to learning situations with a somewhat incomplete and incorrect mental model about the learning topics. Self-explanation can be seen as a process of revising one's existing flawed mental model about the learning materials. This alternative view is introduced below.

The Revision View

The revision view assumes that self-explanation is a process of revising or updating one's imperfect mental model about the learning materials (Chi, 2000). Learners can have unique flaws in what they know about a topic. Therefore, it makes sense that most learners would not generate semantically similar self-explanation inferences for the same piece of missing information, because they customize their self-explanations to fill the gaps in their own mental model. Similarly, it also makes sense that learners do not have a consensus in the loci at which they self-explain because they only need to repair their mental model when they detect a conflict between their mental model and the learning materials. This revision mechanism can further explain why some learners generate more self-explanations, and others explain less in the same vein.

As is discussed previously, conceiving self-explanation as a process of generating inferences will lead to the conclusion that fragmented or incorrect self-explanations could be harmful to learning. However, the evidence failed to support such a conclusion and demonstrated that erroneous self-explanations were harmless to learning (Chi et al., 1989; Chi et al., 1994). From the perspective of revision view, generating incorrect self-explanations might promote learning. In this view, incorrect self-explanations may originate from learners' pre-existing imperfect understanding (mental model) of the topic. The incorrect knowledge generated by self-explaining will then be challenged by the correct information in learning materials, leading to the detection of misunderstandings of the concepts. Such conflicts inevitably elicit further self-explaining episodes of resolving them (Chi et al., 1989). Through a reanalysis of the physics data in the study of Chi et. al. (1989), VanLehn (1999) found that learners were more likely to learn the pieces of knowledge that conflicted with their prior beliefs compared to those consistent with their beliefs, suggesting that incorrect self-explanations actually promote learning.

Accurately detecting conflicts is key to self-repair the imperfect mental model with self-explanation. When self-explaining, learners monitor their comprehension frequently by comparing their mental model with the incoming information. Once a conflict is detected, learners generate self-explanations to revise their incomplete or incorrect understandings about the learning materials, thus improving learning. This is akin to the process of self-reflection (Collins et al., 1988). Therefore, self-monitoring statements can be treated as a high-quality self-explanation. Learners should be encouraged or promoted to self-explain to improve their awareness of self-monitoring.

Self-explanation and Cognitive Load Theory

Cognitive Load Theory (CLT) describes the learning and instructional design implications of a model of human cognitive architecture based on prior knowledge in long-term memory and the information temporarily processed in working memory (Kalyuga, 2011; Sweller et al., 1998). The cognitive load refers to the amount of information that our working memory holds at one time (Sweller, 1988; Sweller et al., 1998). There are three types of cognitive load: intrinsic load, extraneous load, and germane load. Intrinsic load is the working memory load that hinges on the internal task complexity and learners' prior knowledge (Kalyuga, 2009). All other cognitive activities that are not beneficial for learning impose extraneous cognitive load on learners. Germane load is associated with the effort of constructing and automating organized knowledge structures or schemas. Learning methods should increase germane cognitive load and reduce extraneous cognitive load to facilitate learning. Self-explanation is such a constructive learning strategy to hopefully raise germane cognitive load (Renkl & Atkinson, 2003).

Self-explanation enhances germane cognitive load by engaging learners in filling the missing information in learning materials and the gaps in their mental model. These activities make the process of self-explanation time-consuming. Eysink et al. (2009) found, if learners were given enough time, self-explanation-based learning was more effective than hypermedia learning, observational learning, and inquiry learning in terms of learning outcomes and the cognitive load imposed on them. They also found that self-explanation gave rise to the most germane load and caused high overall cognitive load and extraneous load. It has been shown, if the overall cognitive load caused by self-explanation exceeds the working memory capacity, the germane load may become a form of extraneous load, thus impeding learning (Kalyuga, 2009). Therefore, in practice, the extraneous cognitive load caused by self-explanation should be reduced.

Various studies have shown that self-explanation enhanced by prompting can be more effective than spontaneous self-explanation (Atkinson et al., 2003; Chi et al., 1994). It may be the case that prompting learners to self-explain is likely to enhance their germane load in working memory because it forces learners to engage in the construction of relevant knowledge.

Moreover, it appears that self-explanation with some guidance or assistance, e.g., open-ended questions like “Can you explain why problem 1 and 2 are similar?”, are typically more effective than prompts merely saying “Explain this!” (Berthold et al., 2009; Margulieux & Catrambone, 2019; Wylie & Chi, 2014). Eysink et al. (2009) claimed, when learners are not guided or assisted to self-explain, they need to search the relevant subject matter themselves before they can do it. Such searching can easily impose extra extraneous cognitive load on learners and therefore impede learning. Learners guided into self-explanation will reduce the need to search for the relevant subject matter themselves and have more time and capacity available for self-explanation. Consequently, with guidance, the extraneous load imposed on them is reduced and more room is made for germane load in the working memory.

Self-explanation Prompts

Why Prompts Self-explanation?

It is uncommon for learners to spontaneously engage in self-explanation activities while learning (Chi et al., 1989; Renkl, 1997, 2005). Chi et al. (1989) found that only about 10% of learners spontaneously generated self-explanations when they studied worked-out examples about Newton's three laws. Many other studies have replicated this low rate of spontaneous self-explaining (e.g., Hausmann & Chi, 2002; Renkl, 1997, 2005; Renkl et al., 1998). Hausmann and Chi (2002) found that fewer learners generate self-explanations, compared with paraphrasing others' explanations. Renkl and colleagues (2005; 1998) argued that many learners are reluctant to self-explain, especially when they have little prior knowledge because it requires a large amount of effort and mental resources. This deficit of spontaneous self-explanation suggests that self-explaining must be enforced or instructionally assisted (Renkl, 2005).

Definition of Self-explanation Prompts

In self-explanation research, learners are usually encouraged to explain the to-be-learned content by instructional prompts (Bisra et al., 2018; Chi et al., 1994; Renkl, 2005; Stark et al., 1998). Self-explanation prompts are reminders or requests to self-explain the to-be-learned content (Berthold et al., 2009). They elicit self-explanation activities that learners are capable of but do not do so on their own initiative (De Jong & Lazonder, 2014). When prompted to self-explain, most learners can successfully generate and benefit from explanations if they devote additional time and mental resources to the task (Wylie & Chi, 2014).

Prompting Effect

Self-explanation elicited by instructional prompts can also lead to the same learning outcomes as spontaneously generated self-explanation, suggesting that self-explanation itself, whether prompted or intrinsically motivated, benefits learning rather than characteristics of learners who self-explain (e.g., Bielaczyc et al., 1995; Chi et al., 1994; Hausmann & Chi, 2002). Building upon the self-explanation effect, Chi et al. (1994) were the first to investigate whether instructional prompts could elicit self-explanations from learners to help them learn as deeply as good learners who spontaneously generate productive self-explanations. They found that learners who were prompted to self-explain while reading a biology text gained a better understanding of the text passage than the control group who read the passage twice.

A rich body of studies have replicated Chi et al.'s (1994) finding that explicitly prompting learners to generate self-explanations improves learning in various settings (e.g., McNamara, 2004; Renkl, 1997; Renkl et al., 1998; Rittle-Johnson et al., 2017; Schworm & Renkl, 2007). For example, Renkl and colleagues (1998) found that prompting learners to self-explain facilitated them to acquire transferable knowledge. Further investigation revealed that learners with lower levels of prior knowledge benefited more from prompted self-explanations. Griffin et al. (2008) compared three groups of college learners who monitored their understanding levels while reading. One group read a complex text once with the instruction, "Read each text carefully one time, as though studying for an exam." The second group was asked to read the text quickly the first time to get a basic idea of the text, then re-read it more carefully as if studying for a test. The third group was instructed to self-explain while re-reading the text. Results showed that learners who were prompted to self-explain the text outperformed the other two groups to monitor the accuracy of their understanding levels. Therefore, explicitly prompting self-explanation can help learners who do not spontaneously self-explain learn with understanding and foster knowledge transfer (Chi et al., 1994).

Despite its general effectiveness, prompted self-explanation has been observed to have negative or no effects on learning (e.g., Broers & Imbos, 2005; DeCaro & Rittle-Johnson, 2012; Kuhn & Katz, 2009; Mwangi & Sweller, 1998). The mixed results signify that constraints on the effectiveness of prompted self-explanation exist (Rittle-Johnson & Loehr, 2017). For example, Kuhn and Katz (2009) suggested that, under some conditions, prompting to explain one's own solutions or ideas may reduce the effectiveness of self-explanation. Thus, learners may reduce their attention to new information when they repeatedly explain their preexisting mental model. Berthold et al. (2011) found that prompts focused on key concepts increased conceptual comprehension, but reduced transfer. They argued that the conceptually oriented prompts draw learners' limited attention to the key concepts at the expense of neglecting the procedural knowledge. Williams et al. (2013) demonstrated that erroneous overgeneralizations caused by prompted self-explanation could be hazardous to learning. These constraints on the effectiveness of prompted self-explanation appeal to further research to avoid the negative effects of self-explanation but maximize and extend its benefits in practice, e.g., how different types of SE prompts interact with learners' characteristics leads to best learning outcomes.

Types of Prompts

Self-explanation prompts can be classified into different categories based on the amount of guidance they provide and/or the content specificity (Margulieux & Catrambone, 2019). **Content-free or generic prompts** using completely open-ended questions (like "Can you explain this?") provide no guidance at all and avoid reference to the specific learning contents (de Bruin et al., 2007; Hausmann & Chi, 2002). Generic prompts enable learners to express their thoughts non-disruptively and contemplate their understanding (Davis, 2003; King, 1991). **Attention-directed prompts** use focused questions (like “Can you explain how examples 1 and 2 are distinct?”) to direct learners’ attention to the learning content, and thus, provide some guidance (De Koning et al., 2011). **Content-specific prompts** provide the most guidance, referring explicitly to a specific content area and they give learners hints directed toward the solution process in a particular activity, e.g., strategy use (Davis, 2003; Schoenfeld, 2016). Such prompts might draw learners’ attention to difficult conceptual issues that would otherwise be overlooked, but they may also interfere with or override learners’ use of personal standards for metacognitively monitoring content meriting an explanation (Aleven et al., 2006). Additionally, content-specific prompts can elicit self-explanations by many different methods such as open-ended questions (Aleven et al., 2003), selecting explanations from a menu (Conati & VanLehn, 2000), picking problem-solving principles to justify the solution steps (Aleven & Koedinger, 2002), filling in blanks of partial explanations (Berthold et al., 2009). However, content-specific prompts providing too much guidance, like selecting explanations or problem-solving principles from a menu and filling in blanks of partial explanations, may result in active self-explanations rather than constructive self-explanations. According to the ICAP theory, constructive learning is better than active learning (Chi, 2009).

Content-specific prompts may be helpful for some learners to realize that they have gaps in their understanding and even get hints to fill these gaps (VanLehn et al., 1992). However, such prompts may not benefit the learners who already understood these contents. Even worse, they may prevent learners from generating a series of inferences because they direct learners’ attention to specific content (Aleven et al., 2006). On the other hand, Chi (2000) has claimed that generic prompts (e.g., “Explain this sentence to yourself”) should be more effective presumably because they enable learners to tailor their self-explanations for revising their own incomplete or incorrect knowledge structure or mental model. Generic prompts increase the opportunity for learners to detect gaps in their understanding, discover deficiencies in their mental models of the learning contents, or generate useful inferences (Chi, 2000; VanLehn et al., 1992). Given that generic prompts, by definition, are not related to a specific domain they can be implemented in different domains and systems with little to no editing (Kramarski et al., 2013). However, if learners are given too little information, they spend too much of their cognitive capacity trying to figure out what they should learn (Kirschner et al., 2006). For example, Wylie and Chi (2014) found that focused self-explanation prompts, such as "Could you explain how problems 1 and 2 are similar?" were typically more effective than completely open-ended prompts, such as "Could you explain the problems?" They argued that novices know so little about domains that they need clues about what to explain to be most effective.

Different types of self-explanation prompts and different methods of how these prompts are provided seem to result in equal learning outcomes (see Aleven et al., 2006; Aleven et al., 2003). Bisra and colleagues (2018) conducted a meta-analysis investigating the learning outcomes of learners who were prompted to self-explain while studying or solving problems. They found generic (g = .678) and content-specific (g = .510) prompts both produced moderate effect sizes. Though there was no significant difference between the two types of prompts, generic prompts seemed to perform better. Different methods (e.g., interrogative, imperative, fill-in-the-blank, multiple-choice) of how prompts are provided also help learners learn were not found to vary in the effectiveness of improving learning either. Other research showed that attention-directed and content-specific SE prompts are better than content-free/generic SE prompts (Berthold et al., 2009; Gadgil et al., 2012; Johnson & Mayer, 2010). Kwon et al. (2011) found that attention-directed/focused SE prompts had significant advantages over content-specific SE prompts. However, these three types of SE prompts are not directly compared in the same study.

Learner's Apptitudes and SE Prompts

The effectiveness of SE prompts is claimed to depend on learners' expertise (prior knowledge) and learning ability (Aleven et al., 2006; Renkl, 2002; Wylie & Chi, 2014). Renkl (2002) argued that the amount of information provided in the prompt needs to be adapted to the learners' prior knowledge. If given too much information in the prompts, learners with high prior knowledge will be deprived of the opportunity to generate new knowledge by themselves because the information provided in the prompts can fill the gaps in their mental models (Wylie & Chi, 2014). For example, if a SE prompt states "Example 1 and 2 both use the theorem about the probability of the union of events, can you explain how example 1 and 2 are similar?", learners who are knowledgeable of this theorem will not have the opportunity to identify the common structure of the two examples. Whereas, if novices are provided too little information in SE prompts, they have to figure out what information is needed to generate self-explanations by searching their long-term memory and the learning contents, thus increasing the extraneous cognitive load and restraining the capability of self-explanation (Kirschner et al., 2006). For example, if a prompt only states, "Can you explain how examples 1 and 2 are similar?" and novices can hardly recognize the two examples both use the same theorem, they cannot generate further explanations. So, self-explaining is typically more effective for novices when they receive more clues alongside prompts (Berthold et al., 2009; Margulieux & Catrambone, 2019).

Similarly, (Aleven et al., 2006) found that, in ill-defined domains (e.g., legal reasoning), more able learners benefited more from generic prompts (e.g., "Explain this") and less able learners learned better with content-specific prompts which referenced the contents in the transcripts, e.g., “What is the significance of the proximity of the creche to City Hall?”. This is consistent with one of the most common Aptitude-Treatment Interaction (ATI) findings according to Kyllonen and Lajoie (2003): “strong treatments benefited less able learners and weaker treatments benefited more able learners” (p. 82).

However, very few empirical studies have focused on the interaction between learners' aptitude and different types of prompts, which deserves further attention from researchers to make the implementation of the adaptive SE prompts possible in computer-based learning environments, like ITSs.

Computer-supported Prompts

At an early stage of self-explanation research, prompts were usually provided to learners by human tutors or instructors with oral or printed instructions (Chi et al., 1989; Chi et al., 1994). However, computer-supported self-explanation programs, such as ITSs for self-explanation, can make learners more easily access the benefits of self-explanation (Chiu & Chi, 2014). For example, conversational ITSs (e.g. SE-COACH, Conati & VanLehn, 2000), Cognitive Tutor (see Aleven & Koedinger, 2000), AutoTutor (Graesser et al., 2005) can be easily transformed to support self-explanation by interacting with human learners in natural language, enabling learners to get one-on-one and individualized prompts without the presence of human tutors or instructors. Additionally, the interface of an ITS supporting self-explanation enables learners to select explanations from a menu, type, or speak aloud to generate self-explanations which will then be kept in the system in text or voice recording formats. Also, ITSs can decide when and how the prompts should be provided based on the learner model (Conati & VanLehn, 2000).

As early as 2000, Conati and VanLehn have developed the SE-COACH, an ITS that monitors and supports self-explanation on worked-out examples in the domain of physics. The prompts designed in SE-COACH to elicit self-explanations are either through stimulating self-questioning (e.g., “the choice is correct because...”) or through selecting explanations from a menu to justify the steps. Koedinger and colleagues (Koedinger, Corbett, Ritter, & Shapiro, 2000) developed a new version of Geometry Cognitive Tutor that supports self-explanation. In Geometry Cognitive Tutor, learners are prompted to explain their own solution steps by naming the problem-solving principles that justify their steps. Aleven et al. (2001; 2002) even developed a tutorial dialogue system to elicit learners to generate self-explanations.

In the current study, AutoTutor, an intelligent tutoring system that helps learners achieve deep levels of learning by holding conversations in natural language, is used to prompt self-explanations (Nye, Graesser, et al., 2014). In AutoTutor, the computer agents can be used to elicit self-explanations by talking to the learner and displaying the instructional prompts in its interface. The learners can type their explanations in a dialogue box. Schworm and Renkl (2006) found that written self-explanations are better than spoken explanations because they require articulating thoughts and creating a record, which allows learners to reflect on their explanations more easily.

Research has shown that learners can also benefit from self-explanation prompts provided by computers (Aleven & Koedinger, 2002; Hausmann & Chi, 2002). Aleven and Koedinger (2002) demonstrated that the benefits of self-explanation could be achieved when learning with the Geometry Cognitive Tutor that supports self-explanation. They found that learners who were prompted to self-explain their own solution steps learned with greater understanding compared to learners who did not explain their steps. Atkinson et al. (2003) showed that prompting principle-based self-explanations in a computer-based learning environment that provided worked-out examples led to superior learning outcomes in terms of performance on similar problems and novel problems in the domain of probability. Conati and VanLehn (2000) as well as Schworm and Renkl (2006, 2007) provided further evidence for the positive effects of self-explanation prompting when learning from computer-based worked-out examples was provided. The results from these studies demonstrate that computer-supported prompts can exert their effects on the benefits of self-explanation. Moreover, once the computer-supported prompts, especially prompts provided by ITSs, are designed, they can be easily reused and have additive benefits on learning. For example, prompts can be provided adaptively (when self-explanations should be scaffolded and when they should fade) based on learners' gaps in understanding or expertise (Leppink et al., 2012).

In summary, self-explanation usually needs to be prompted to play its role in constructing learning materials to help learners gain a better understanding and apply new knowledge outside the corresponding contexts. However, constraints may exist when different types of prompts are applied to learners with various aptitudes. Further research is needed to figure out how to make the best use of SE prompts, especially in computer-based environments, like ITSs.

Self-explaining Worked Examples

Learning from worked examples is an effective and efficient way for learners to improve their learning and facilitate knowledge transfer. Greater learning gains and deeper understanding can be achieved with less investment of time and mental efforts (Atkinson et al., 2000; Sweller et al., 1998; Van Gog et al., 2011). This is known as the "Worked Example Effect". However, it is argued that, although learning from worked examples may result in initial performance improvement for learners, it does not promote deep processing of concepts because learners usually prefer to rely on analogical mapping of example steps to problem steps, rather than expanding their cognitive efforts to make inferences about the concepts and generalities from the examples (Eiriksdottir & Catrambone, 2011). The analogical mapping may aid near transfer but not far transfer, that is, learners can apply the concepts in the examples to similar problems, but they cannot solve problems that require slight deviations from the examples (Sweller & Cooper, 1985). Self-explaining worked examples can remedy this deficit and further improves learning processes by integrating learners' relevant prior knowledge and new information to explore the plausible explanations of the worked example steps (Chi et al., 1994; Sweller, 2010). These processes help learners build a better mental representation of the problem-solving procedures that allows them to apply their knowledge more easily to novel problems (Renkl & Atkinson, 2003).

Worked examples are "models of correct behaviors", that is, they are all correct examples by default. It has been established that explaining correct examples increases learners' conceptual knowledge (Hilbert et al., 2008) and their ability of both near and far transfer, and the ability to solve both similar and more difficult problems (Renkl et al., 1998). However, Siegler (2002) demonstrated that, for children learning math, explaining correct and incorrect solutions was better to solve transfer problems than to explain correct solutions only. Other research showed that explaining both correct and incorrect examples can further improve conceptual understanding, procedural skills (Booth et al., 2015; Booth et al., 2013), and reduce misconceptions (Durkin & Rittle-Johnson, 2012). Compared to explaining only correct examples, explaining incorrect examples, no matter they are paired with correct examples or not, can lead to a greater conceptual understanding of the learning content (Booth et al., 2013). Explaining errors can draw learners' attention to the specific features in a problem that make the procedure inappropriate. This can help the learner replace faulty conceptual knowledge they have about the meaning of the problem features with correct conceptual knowledge about those features; the acquisition of accurate, deep features with which to represent problem situations is key to building expertise (Chi et al., 1981). However, Booth et al. (2013) claimed that they did not support that the learners should only explain incorrect examples with the reason that correct worked-out examples help learners build correct knowledge unless they could get correct knowledge from other sources, e.g., further practice with feedback, or re-read the textbook.

The current study will build on the evidence above and ask learners to explain both correct and incorrect solutions. Thus, the effort will be only devoted to exploring the interaction between different types of SE prompts and learners' aptitudes.

The Current Study

The current study explored the effects of three types of SE prompts on learners with different levels of prior knowledge and learning ability when learning *Probability theory* in a conversation-based ITS - AutoTutor. AutoTutor is a conversation based ITS which supports a mixture of vicarious learning and interactive tutoring (Nye, Graesser, et al., 2014). In vicarious learning, human learners learn by observing a tutoring interaction between one or more animated peer agents and the tutor agent. Here, the peer agents ask deep questions which are promptly answered by the tutor agent. In interactive tutoring, human learners answer a main question by interacting with a tutor agent in natural language. AutoTutor provides an ideal experimental environment for learning science research. AutoTutor can be used to display learning materials (e.g., Google slides, images, or text). Additionally, conversational computer agents can provide experimental instructions and guide learners through the experiments. Two topics of probability theory extracted from a textbook (see Hogg et al., 2010) were used as the learning content. One topic was the Properties of Probability, which involves basic knowledge of Set Theory, Venn diagram, the definition of events, properties of Set operations, the definition of probability, and some theorems of probability. The other topic was enumeration which includes the multiplication principle, permutation, and sampling. These two topics are the first two topics in the first chapter of the textbook and can be learned by anyone who has basic knowledge of mathematics.

The procedure of the current study started with learners receiving didactic lessons on one of the two topics so that they could acquire the intended knowledge. Afterward, learners' prior knowledge and learning ability were assessed by a test consisting of conceptual problems and procedural problems. Then, learners explained the procedural problems they had worked on in the pretest phase. Finally, learners received a posttest that was similar to the pretest but with both near and far transfer tests. The near transfer test included isomorphs of the previously explained worked-out problems. The far transfer test included problems that were related but not isomorphic to the self-explained problems, that is, they were different problems but can be solved using the same theorems or principles as the worked examples (Haskell, 2001).

Content-specific, Generic, and Guided Prompts

The three types of SE prompts are content-specific prompts, generic prompts, and generic prompts with a form of guidance. Content-specific prompts provided the most guidance and generic prompts provided the least guidance. The generic prompts with a form of guidance will be called guided prompts for short. These prompts are implemented using open-ended questions based on two considerations. First, according to Davis (2003) and King (1991), the open-ended nature provides explainers a non-disruptive opportunity to express their thoughts and contemplate their understanding. Second, articulating self-explanations in their own language help learners improve mathematical communication which is considered an essential ability for twenty-first-century learners learning mathematics, according to the National Council of

Teachers of Mathematics (NCTM, 2000). Content-specific prompts elicit learners to explain why a problem-solving principle has been misused in the incorrect solutions, like “Why is the probability of event A intersecting with event B,0?”. Generic or content-free prompts merely ask learners to “explain why this solution is incorrect.” The guided prompts are an invention of the current study. They are the same as the generic prompts in form, but learners will be provided with some guidance about what errors learners may commit during solving probability problems. O'Connell (1999) identified four categories of errors learners might commit at different points in *Probability* problem-solving. They are text comprehension errors, conceptual errors, procedural errors, and arithmetic errors (see Table 1). These errors can be taken as background meta-knowledge which is delivered to learners as guidance before they explain the incorrect solutions. They also serve as tips for learners to consider the correct solutions from these four aspects. These three types of SE prompts compose the three conditions of the experiment in this study.

Table 1. *Error Categories (cited from O'Connell, 1999)*

Category	Description
Text Comprehension	General misunderstanding of the information contained in the text of a problem, such as assigning a probability value to the wrong event, incorrectly identifying the goal of a problem, misinterpreting statements involving inequalities, etc.
Conceptual	Errors involving basic concepts or definitions of probability, such as reporting a negative probability value or a probability greater than 1.0, assuming events are equally likely without appropriate justification, applying the algebra of real numbers to sets, equating frequency with probability, misunderstandings of independence, mutually exclusive events, or complementary events, etc.
Procedural	Faulty procedures, such as: forgetting outcomes when defining a sample space, not checking preconditions before applying a formula (i.e., for ME or Independent events); using an incorrect version of a formula, forgetting values or substituting incorrect values into a formula, inventing incorrect procedures, using inappropriate strategies, or not completing a strategy, substituting the wrong values into an expression, etc.
Arithmetic	These are errors involving simple miscalculations, copy mistakes such as transposing digits, incorrect cancellation of terms from numerator and denominator of an expression, etc.

The Quality of Self-explanations

Schworm and Renkl (2006) have shown that the quality of written self-explanations is a predictor of learning outcomes. As mentioned above, high-quality self-explanations include SE inferences and statements of metacognitive monitoring (McNamara, 2004). However, these types of self-explanations are extracted from protocols of text and comprehension experiments. The current study will adopt the categories of high-quality self-explanations proposed by Berthold et al. (2009) as well as their coding method because they also studied the self-explanation effect on the domain of probability. The two categories of high-quality self-explanations are:

i). Elaboration-based self-explanations: this category includes elaborated principle-based self-explanations and elaboration of errors. Learners usually self-explain a solution by identifying the underlying domain principles. However, if a principle is merely mentioned without being elaborated (e.g., "mutually exclusive") in a piece of self-explanation, this category will not be scored. There should be some elaboration when referring to a principle (e.g., "the two events are mutually exclusive, so the probability of the intersection of two events is 0"). When justifying incorrect solutions, learners should elaborate what the errors are (e.g., "the learner did not recognize that the two events are mutually exclusive", or "either event A or B happen is to calculate the union, not the intersection, of them,") based on the four categories of errors (O'Connell, 1999).

ii). Rationale-based self-explanations: this category refers to high-quality self-explanations that include the rationale of the domain principles and why incorrect solutions are wrong. The rationale-based self-explanations give reasons for why the principle is as it is, not just elaborate it or state the correct application conditions of the principle. For example, when the theorem about the probability of the union of two events A and B ($P(A \cup B) = P(A) + P(B) -$

$P(A \cap B)$) is used to solve a problem, a learner who elaborates it will say "the probability of the union of events A and B equals to the probability of A plus probability of B minus the probability of the intersection of A and B"; a learner who states the application conditions of the theorem will say "if A and B are not mutually exclusive events, we can apply it in this problem"; while the rationale-based self-explanation will be "the probability of the union of events A and B equals to the probability of A plus probability of B minus the probability of the intersection of A and B, this is because both event A and B contain the intersection of A and B, thus the intersection will be repeated twice when computing A plus B, so we have to get rid of one of them." When justifying incorrect solutions, learners should give the reason why the errors lead the solutions to be wrong and how they can be corrected (e.g., "the mutually exclusive events do not intersect with each other. They cannot happen at the same time. So, the probability of the intersection of them is 0.", or "either/or means no matter which event of the two happens is acceptable, both means both events should happen at the same time, so we have to calculate the probability of union.").

Three experts on probability coded the self-explanations. They were blind to the experimental conditions. The three coders were trained to identify the two categories of self-explanations by the experimenter. The inter-rater reliability was calculated to measure the degree of agreement among them. In cases of divergence, the final coding was determined by discussion.

Prior Knowledge and Learning Ability

Declarative knowledge and procedural knowledge were assessed as learning outcomes. Declarative knowledge includes the facts, concepts, principles that apply within a domain (De Jong & Ferguson-Hessler, 1996). Procedural knowledge is often defined as knowledge of procedures (Rittle-Johnson et al., 2001; Star, 2005, 2007). A procedure is a series of steps, or actions, done to accomplish a goal. This knowledge often develops through problem-solving practice, and thus is tied to particular problem types. Procedural knowledge was assessed by learners' problem-solving performance after the knowledge acquisition phase. Prior knowledge is defined as the amount of declarative knowledge the learners obtain after learning the text about two topics of probability at the knowledge acquisition phase. Learning ability was assessed by the amount of declarative knowledge the learners recalled combined with their problem-solving performance. Learning ability involves the memory process, understanding of the learning content, and the ability to transfer the knowledge to new contexts. It conforms with the memory and implicit learning processes described by Woltz (2018). The memory processes involve recall, recognition, and implicit memory processes that are revealed in performance facilitation, often without content-specific retrieval intent by the learner and despite lack of awareness of the original learning event or events. The implicit learning processes involve learning new procedures to solve problems and invoking new ideas based on the learned items.

Procedural Transfer

Procedural transfer is the adaptation and/or integration of procedures to solve problems with structural and surface features that differ from the learning phase (e.g., requiring the use of learned procedures in new combinations) (Atkinson et al., 2003; Wong et al., 2002). Procedural transfer involves two kinds of transfer: near transfer and far transfer. Near transfer test includes problems that are isomorphic but not identical to worked-out problems that are explained; that is, they are the same problems but with different parameters or scales of number. The far transfer test includes problems that are related but not isomorphic to the self-explained problems; that is, they are different problems, but they can be solved using the same theorems or principles as the worked examples (Haskell, 2001).

Research Questions and Hypotheses

First**,** based on the literature reviewed, computer-supported SE prompts attained positive effects on the learning outcomes (Atkinson et al., 2003; Conati & VanLehn, 2000; Schworm & Renkl, 2006, 2007). Also, explaining both correct and incorrect examples benefits conceptual understanding and knowledge transfer more than only explaining correct versus incorrect examples (Booth et al., 2015; Booth et al., 2013; Durkin & Rittle-Johnson, 2012; Siegler, 2002). These led to the first hypothesis that the three types of SE prompts (generic, guided, and content-specific) implemented by AutoTutor are generally effective in improving learning (RQ 1).

- Hypothesis H1: The posttest response accuracies of all learners of the three conditions (generic prompts, guided prompts, and content-specific) are higher than existing pretest response accuracies.

The second research question addresses which SE prompts are more effective in improving learning outcomes in general. Chi (2000) has claimed that generic prompts (e.g., “Explain this to yourself”) should be more effective than content-specific prompts presumably because they enable learners to tailor their self-explanations for revising their own incomplete or incorrect knowledge structure or mental model. Generic prompts increase the opportunity for learners to detect gaps in their own understanding, discover deficiencies in their mental models of the learning contents, and generate useful inferences (Chi, 2000; VanLehn et al., 1992). Moreover, content-specific prompts may not benefit the learners who already understood these contents, or even worse, they may prevent learners from generating a series of inferences because they direct learners’ attention to specific content (Aleven et al., 2006). However, if learners are given too little information, they spend too much of their cognitive capacity trying to figure out what they should explain (Kirschner et al., 2006). For example, Wylie and Chi (2014) found that focused self-explanation prompts, such as “Could you explain how problems 1 and 2 are similar?” were typically more effective than completely open-ended prompts, such as “Could you explain the problems?” They argued that low prior knowledge learners know so little about a domain that they need support to make self-explanation more effective.

Sometimes too much support will turn self-explanation into forms of active learning that reduce its benefits. The guided prompts possess the advantages of generic prompts and provide learners with a form of guidance that does not direct learners’ attention to specific content. The guidance refers to the meta-knowledge of errors that learners may commit during probability problem-solving. At different steps of problem-solving, learners may ask themselves “Is the solution based on the correct comprehension of the question?”, “What are the theorems or concepts applied here?”, “What is the procedure of the solution?” Such meta-knowledge will reduce the cognitive load of learners when they search for knowledge to explain the solutions. As a result, learners benefit more from the generic SE prompts. These theoretical implications and empirical evidence lead to the following two hypotheses:

- Hypothesis H2a (guided > content-specific > generic):
    1) Learning gains of the learners in the Guided condition are greater than the other two conditions (generic and content-specific).

       Learning gains of learners in the content-specific condition are greater than those who are in the generic condition.
- Hypothesis H2b (guided > content-specific > generic):

1) Learners in guided condition generate more elaboration-based and rationale-based self-explanations than those in content-specific condition.
2) learners in the content-specific condition generate more elaboration-based and rationale-based self-explanations than the generic condition.

Third, from the available literature review, we observed that interactions might exist between SE prompts and learners' aptitude (prior knowledge and learning ability) (Aleven et al., 2006; Renkl, 2002; Wylie & Chi, 2014). Renkl (2002) argued that the amount of information provided in the prompt needs to be adapted to the learners' prior knowledge. If given too much information in the prompts, learners with high prior knowledge will be deprived of the opportunity to generate new knowledge by themselves because the information provided in the prompts is already capable of filling the gaps in their mental models (Wylie & Chi, 2014). Other researchers found that learners with low prior knowledge typically benefit from prompts with more clues or information (Berthold et al., 2009; Margulieux & Catrambone, 2019). However, according to Wylie and Chi (2014), SE prompts with some support have more advantages in learning compared to with no support at all. So, the hypotheses below were tested and compared:

- Hypothesis H3a:
    1) Learners with high prior knowledge benefit most from guided SE prompts in learning gains than from generic SE prompts, and least from content-specific SE prompts (guided > generic > content-specific).
    2) Learners with low prior knowledge benefit most from content-specific SE prompts in terms of learning gains than from guided SE prompts, and least from generic SE prompts (content-specific > guided > generic).

With respect to learners’ learning ability, Kyllonen and Lajoie (2003) claim that “strong treatments benefited less able learners and weaker treatments benefited more able learners” (p. 82). According to Cronbach and Snow (1977), the strong treatments here imply the self-explanation prompts with more guidance, and weak treatments imply the prompts with less or no guidance. In this study, content-specific prompts provided the most guidance. They could be seen as strong treatments. Guided prompts provided less guidance compared to content-specific prompts, so that they could be seen as weaker treatments. Furthermore, the generic prompts provided no guidance at all. They could be seen as the weakest treatments. The less able learners refer to learners with lower learning ability, and more able learners are learners with higher learning ability (see the definition in section *Prior Knowledge and Learning Ability*). Aleven et al. (2006) found that less able learners benefit more from content-specific self-explanation SE prompts, and more able learners benefit more from generic self-explanation prompts. So, we hypothesize that:

- Hypothesis H3b:
    1) High-ability learners benefit most from generic SE prompts in learning gains than from guided SE prompts, and least from content-specific SE prompts (generic > guided > content-specific).
    2) Low-ability learners benefit most from content-specific SE prompts in learning gains than from guided SE prompts, and least from generic SE prompts (content-specific > guided > generic).

The fourth research question this study explored was whether high-quality self-explanation could predict far procedural transfer. As we know, self-explaining supports both comprehension and far transfer of knowledge (Berthold et al., 2009; Chi et al., 1989; Rittle-

Johnson & Loehr, 2017). The essence of self-explanation lies in generating inferences or new knowledge by integrating learners' prior knowledge and current learning materials (Chi, 2000). This implies that a complete mental model facilitates knowledge transfer. Chi (2000) defined self-explanations strictly as inferences that refer to high-quality self-explanations. The study of Chi et al. (1989) indicated that high-quality self-explanations generated by learners are positively related to knowledge transfer. This leads to the last hypothesis:

- Hypothesis H4: High-quality self-explanations predict the far transfer of problem-solving procedures.

Methods

Learners

Learners of the current study were recruited from an online crowdsourcing platform, Amazon Mechanical Turk (AMT). Three hundred and sixty-five learners were recruited on AMT. However, only 129 learners completed the entire study and self-explained the worked-out examples. Of these 129, 35 (27.1%) were female, and 94 (72.9%) males. Their ages ranged from 21 to 60 with a mean of 33.8 years old. All the available learners had at least a high school degree or equivalent, among whom 4.7% learners had a high school degree or equivalent, 65.9% learners had a bachelor's degree, 23.3% learners had a master's degree, 3.9% learners had a doctorate degree, and 2.3% learners chose "Other" and specified them as mathematics professionals. One question inquired how much knowledge a learner has about Probability. The levels included "not at all", "a little", "moderate amount", "a lot", and "professional." The results revealed that 6.2% of the learners reported that they were completely unfamiliar with probability, 23.3% learners had "a little" knowledge about probability, 27.1% learners had a moderate amount of knowledge, 27.1% learners had "a lot" of knowledge, and 16.3% learners considered themselves professionals in the area. The learners were randomly assigned to the three conditions, but the numbers of learners regarded as usable were 36, 48, and 45 in the generic, guided, and Content-specific conditions, respectively.

Materials

The experiment was integrated with a Qualtrics survey and conducted on Amazon Mechanical Turk (AMT). The experiment involved the demographic survey, learning materials to read, pretests, AutoTutor sessions for self-explanation, and posttests. The details of the implementation of these phases are discussed below.

Demographic Survey

The demographic survey contained four questions about the learners' age, gender, the highest degrees or levels of school they have completed, and how much knowledge they have about probability (see Appendix II).

Probability Topics

The learning content of the current study included two topics about probability theory that were extracted from a widely used textbook *Probability and Statistical Inference, 9th Edition* (see Hogg et al., 2010) in college level Probability and Statistics courses. Topic I was the properties of probability (PP) which involves basic knowledge of Set Theory, Venn diagrams, the definition of events, properties of Set operations, the definition of probability, and some theorems of probability. Topic II was on methods of enumeration (ME), including multiplication principle, permutation, and definitions about sampling. These two topics are simplified versions of the first two sections (*1.1 Properties and Probability*, and *1.2 Methods of Enumeration*) of the first chapter (https://www.pearson.com/us/higher-education/program/Hogg-Probability-and-Statistical-Inference-9th-Edition/PGM91556.html).

The reading materials of the two topics were organized in the form of concepts and theorems mixed with examples to illustrate their application. The texts of the two topics were converted into presentation slides (see Appendix I). Meanwhile, to better explicate the concepts and theorems, some figures were added to the slides to make the examples more vivid. For example, in the example to illustrate the use of the multiplication principle in topic ME, the figures of male and female rats and different types of drugs were added in the tree diagram (see Figure 2).

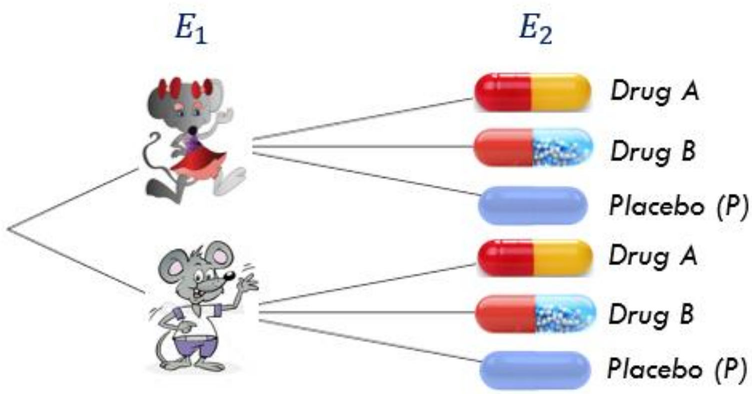


Figure 1. *Tree Diagram*

AutoTutor

AutoTutor is a conversation-based Intelligent Tutoring System that supports a mixture of vicarious learning and interactive tutoring (Nye, Graesser, et al., 2014). Vicarious learning is implemented in the *Information Delivery* mode. Human learners learn from the didactic information provided by the tutor agent and the observation of the tutor agent promptly answering deep questions asked by peer agents. In interactive tutoring mode, human learners answer a main question by interacting with the tutor agent in natural language. The dialogue pattern during this process is called *expectation and misconception tailored dialogue* (EMT dialogue) which can be commonly observed from the interactions between human tutors and learners (Graesser et al., 1995). The EMT dialogue is the primary pedagogical method that attempts to implement explanation-based learning (Graesser, 2016; Graesser et al., 2004). Each main question is associated with a list of expectations (anticipated good answers, steps in a procedure) and a list of anticipated misconceptions (bad answers, incorrect beliefs, errors, bugs). As the learners express their answers over multiple conversational turns, the information they provide is compared with the expectations and misconceptions. AutoTutor gives positive (e.g., “Great answer”), neutral (e.g., “I see”, “Uh huh!”), or negative (e.g., “Not really”, “Not quite”) feedback to the learner based on the quality of the answers, pumps the learner for more information (e.g., with the question “What else?”), prompts the learner to fill in missing words, gives hints to direct the learner to answer the main question, fills in missing information with assertions, identifies and corrects bad answers, answers learners’ questions, and summarizes answers at the end of dialogue turns.

In the proposed study, both the information delivery mode and the interactive tutoring mode were used but not in the traditional way described above. The information delivery mode was used to present the instruction and training materials of the self-explanation sessions. The interactive tutoring mode was used as a learning environment for learners to self-explain the worked examples. The main questions were replaced by SE prompts spoken by a talking head and displayed in the interface. Learners typed their self-explanations in the textbox after being prompted.

Self-explanation sessions for the two topics were implemented as two separate AutoTutor technical components (Shareable Knowledge Objects [SKOs], Nye, Rahman, et al., 2014). The SKOs can be seen as lectures or lessons that are delivered by web pages (see Appendix I). In the SKOs for self-explanation, the web pages display four computer agents, the worked examples and their correct or incorrect solutions, as well as textboxes for learners to type their explanations. The four computer agents interact with learners by prompting them to explain the solutions to the worked examples. All learners' responses in AutoTutor were sent to and stored in a learning record store that uses a standard (xAPI, Kevan & Ryan, 2016) to format the data.

Figures 2 and 3 below illustrate the interactions between the learners and AutoTutor agents. In the top right corner, the avatar is a teacher agent, Ben, who gives instructions on what the learners are expected to do and presents the correct solutions. For example, the teacher agent would say:

> "Please read the correct solution to the question in the center of the page and explain why this solution is correct. Then type your explanations in the pop-up window. After you finish typing, click the "crossing" button in the pop-up window to close the window and submit your explanations."

The other three avatars are learner agents. They are Angela, Anna, and Carl from top to bottom. Each of them presents a wrong solution to learners. They claim the wrong solutions were their solutions and ask the learners to figure out why the solutions are wrong and explain to them. Then, learners read the solutions, try to figure out whether they understand the solution, type their explanations in the pop-up window (Figure 3), and submit them (which gets recorded in a learning record store that tracks all of the actions of the learner). The learners do not get any feedback from the agent about the quality of their explanations.

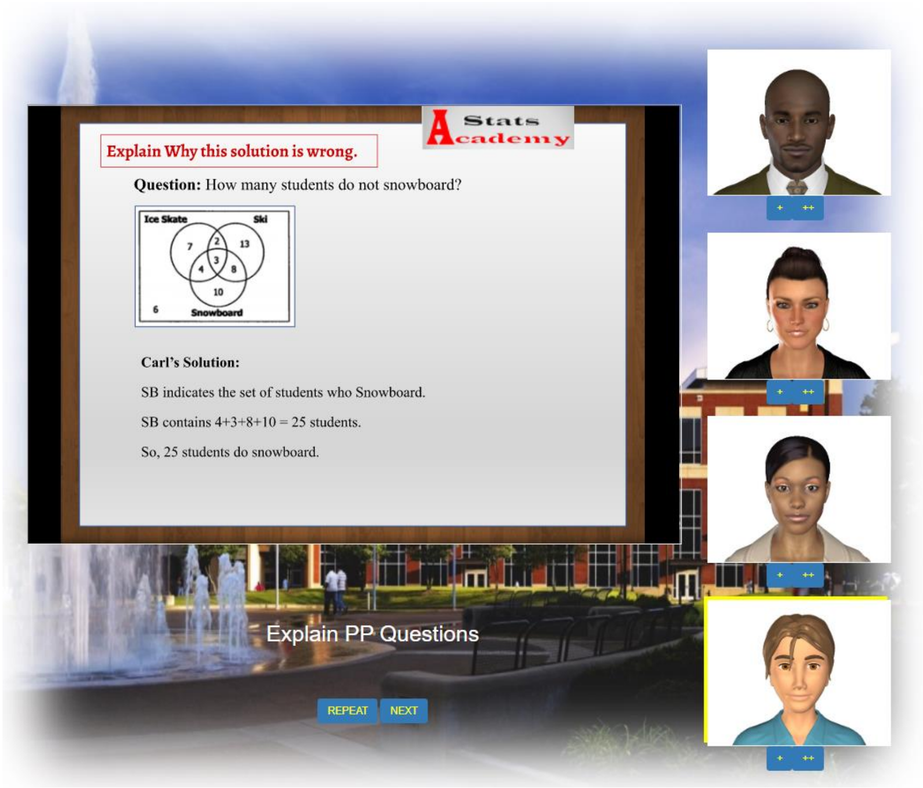


Figure 2. *Self-explanation in AutoTutor*

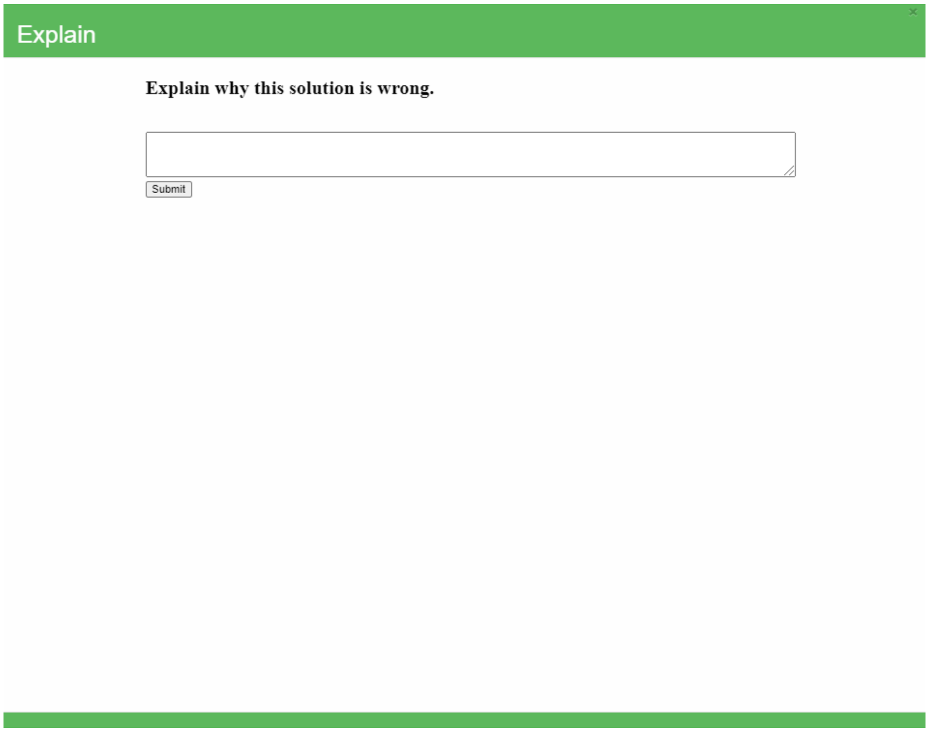


Figure 3. *Pop-up Window for Typing Self-explanations*

Qualtrics

Qualtrics is an online survey development environment. It was used to distribute the learning materials and administer the demographic survey, the pretests, and the posttests. It also provides considerable flexibility for advanced users by enabling them to develop their surveys using JavaScript and HTML. In the proposed study, the URLs of the self-explanation sessions were embedded in the Qualtrics survey using HTML and JavaScript. The slides of the reading materials were broken into separate pictures and embedded each slide in a block with a Timing module in the survey. By doing so, the time learners spent on each of the slides was recorded. As we previously stated, self-explanation interventions were implemented in AutoTutor.

One problem was that the data generated in Qualtrics and AutoTutor were stored separately. A solution was needed to match learners' identities in the two databases. Upon observing the data generated in Qualtrics, it was apparent that Qualtrics uses a unique "responseID" to identify a specific user. Using JavaScript, the unique "responseID" of a learner can be passed to the learning record store of AutoTutor when they click on the AutoTutor link. In addition, all of the questions in the pretests and posttests were timed with the Timing module in Qualtrics.

Pretests and Posttests

The conceptual and procedural problems in the pretest and posttest were based on the contents of the two topics (PP and ME) and collected from the textbook *Probability and Statistical Inference, 9th Edition* (see Hogg et al., 2010) and online learning websites. There were two versions of tests for each topic. Each test for the topic, *properties of probability*, contained 13 conceptual questions and 12 procedural questions; and each test for the topic, enumeration methods, contained 6 conceptual questions and 14 procedural questions (see

Appendix III). The two versions of tests were given to learners either as a pretest or as a posttest for counterbalancing. In each of the three conditions (Content-specific, Generic, and Guided), learners were randomly selected to receive one version of the tests as the pretest and the other version as the posttest. The posttest for each topic contained the same questions as those in the pretest but with different parameters.

To avoid experimenters' bias, four procedural questions in the pretests of the two topics were randomly selected for learners to explain. The rest of the questions (8 for Topic I, 10 for Topic II) in the pretests were not explained. The counterparts of the unexplained questions in posttests served as the **far transfer tests**. Each to-be-explained question had a correct solution and three incorrect solutions. Learners explained why the correct solution was right, and why the incorrect solutions were wrong.

Procedure

Figure 5 illustrates the procedure of the experiments. After giving informed consent to participate, AMT redirects learners to the Qualtrics survey of the experiment hosted on https://memphis.co1.qualtrics.com/. After reading the instructions for the experiment, learners answered demographic questions inquiring about their age, gender, the highest degrees or levels of school they have completed, and how much knowledge they have about probability (See Appendix II).

The two probability topics were studied in the same four steps. The topics were provided to learners in random order. In the first step, learners learned a topic by reading the learning material at their own pace in Qualtrics. Time spent on each slide was recorded and saved in Qualtrics. In step 2, the learners' prior knowledge and learning ability were assessed by one version of the tests for the two topics (see Appendix III). learners' accuracy and time on each question of the pretest were collected. In step 3, they were asked to explain the correct and incorrect solutions to four of the procedural problems. learners' self-explanations were collected. Feedback to indicate whether the explanations were good or bad was not given to learners. This was because feedback is a form of instruction. If learners were given feedback, they would learn from it. As a result, it would be unclear whether learners' learning was from self-explanation or feedback. In the fourth step, learners were administered a posttest (see Appendix III).

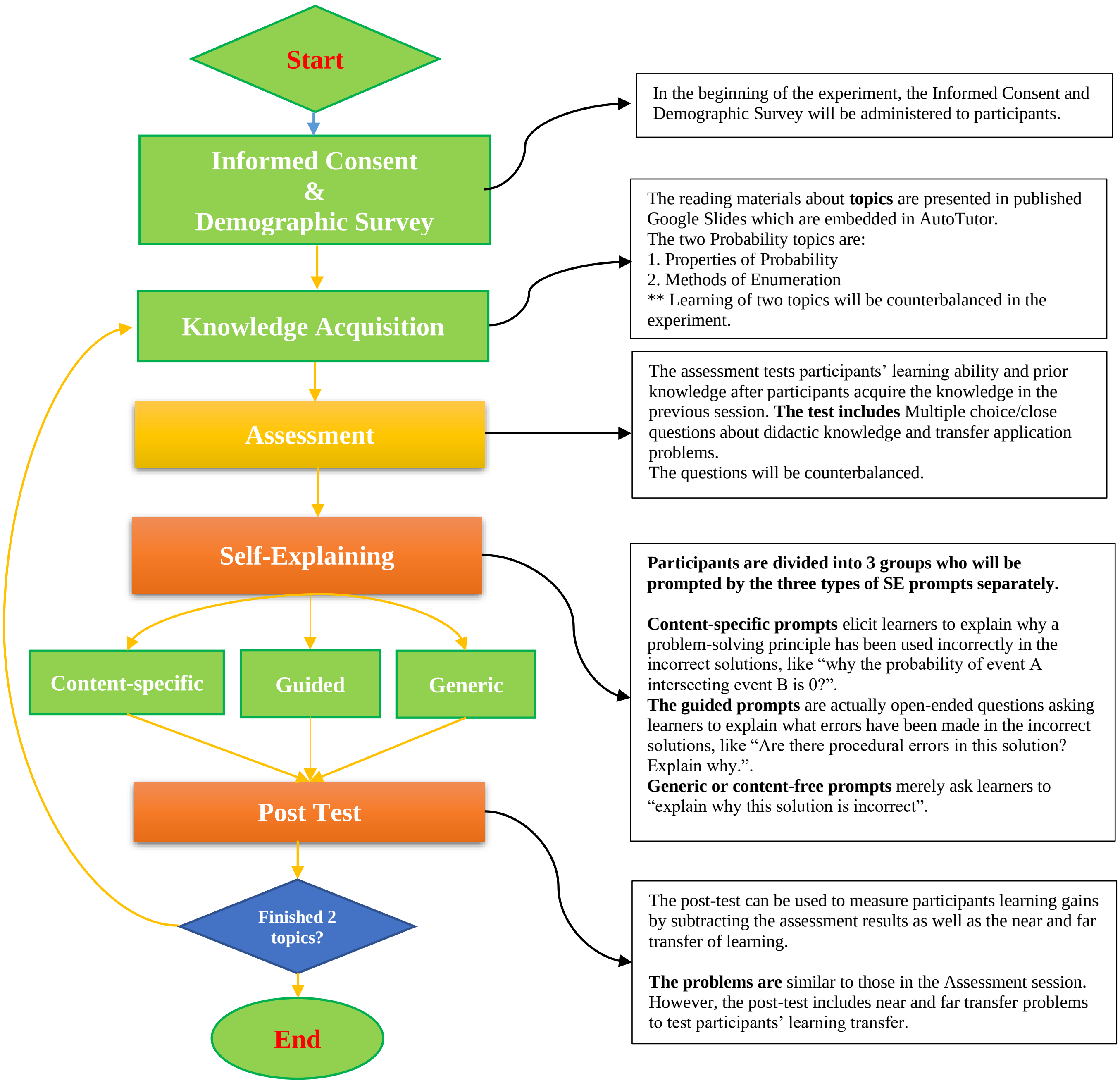


Figure 4. *Design of the Experiment*

Measures

In the pretests and posttests, learners' performance was measured by the score they received from answering the questions. Learners received a score of 1 for each question if they correctly answered the question, 0 otherwise. The learning gains were measured by the normalized change (c) which involves the ratio of the gain to the maximum possible gain or the loss to the maximum possible loss (Marx & Cummings, 2007). If a learner's performance improved from the pretest to the posttest, we used the equation $(post - pre)/(1 - pre)$ to calculate this learner's normalized change, where pre and post are the pretest and posttest scores out of 100, respectively. If a learner's performance worsened, we used the equation $(post - pre)/pre$, which is the ratio of the actual loss to the maximum possible loss. If the learner's pretest score was equal to the posttest score, c = 0. Learners who earned a perfect score on the pretest and posttest were removed from the data. Likewise, learners who scored 0 on both the pretest and posttest were removed from the data sets. A summary of these quantitative possibilities is summarized in equation 1.

$$c = \begin{cases} \dfrac{post - pre}{100 - pre}, & post > pre \\ drop, & post = pre = 100 \; or \; 0 \\ 0, & post = pre \\ \dfrac{post - pre}{pre}, & post < pre \end{cases} \tag{1}$$

where post and pre refer to the posttest and pretest scores out of 100, respectively.

The amount of time a learner read a slide or answered a question was measured by the duration (in seconds) from the onset of a slide or question page to when they left the page.

The learners' prior knowledge was measured by the proportions of their scores on conceptual questions in the pretests. Learners' learning ability was measured by the proportions of their total scores on the pretests. There were 13 conceptual questions and 12 procedural questions in the pretest of Topic I, and 6 conceptual questions and 14 procedural questions in the pretest of Topic II. All of them were multiple-choice questions. Learners received a score of 1 if they answered a question correctly; 0 otherwise. So, learners' scores on conceptual questions in the pretests of Topic I and Topic II ranged from 0 to 13 and from 0 to 6, respectively. Their total scores on the pretests of Topic I and Topic II ranged from 0 to 25 and from 0 to 20, respectively. The proportions of these scores were calculated by the ratio of these scores to the possible maximum scores the learners could receive.

Four procedural problems were randomly selected from the pretest of each topic for learners to explain. Therefore, eight questions in the pretest of Topic I and ten questions in the pretest of Topic II were not explained. These questions served as pretests of far transfer tests for the two topics. Their counterparts in the posttests served as posttests of far transfer tests. The far transfer of problem-solving procedures was measured by the normalized change of learners' scores from the far transfer pretest to the far transfer posttest.

Learners' self-explanations on each solution (either correct or incorrect) were rated by three graduate learners who were experts on probability. The three raters were trained to identify high-quality self-explanations by reading the *The Quality of Self-explanation* section, which the experimenter also explained. Self-explanations of a solution were given a score of 1 when a rater considered them as high quality, otherwise 0. The discrepancy between raters was resolved and the final ratings were generated by the mechanism that, if more than two raters considered the self-explanations of a solution as high quality, they were given a score of 1, otherwise 0. The quality of self-explanations was measured by the proportion of high-quality self-explanations (see *The Quality of Self-explanations* section) which was the total score a learner received from the final ratings divided by the total number of solutions that were explained. The total number of solutions to be explained was 16 for both topics.

Data Analyses

To test hypothesis H1, three single sample t-tests were performed to examine whether learners' learning gains (normalized change) were significantly greater than 0. Hypothesis H1 that three types of SE prompts (content-specific, generic, and guided) improve learning will be confirmed if learners' learning gains (normalized change) of the three conditions are greater than 0.

To test hypotheses H2a/b, multiple linear models were performed using *lm* package in R to compare the difference of learning gains and the numbers of high-quality self-explanations between three conditions of SE prompts (content-specific, generic, and guided) (Chambers et al., 1992; Wilkinson & Rogers, 1973). Individual difference variables including age, gender, educational levels, and the self-reported knowledge level on probability were added in the models as control variables. In the regression models, Male was the baseline for Gender. Education levels and self-reported knowledge levels of probability were taken as ordinal variables. The content-specific condition was the baseline for the comparison of the three conditions of self-explanation prompts. For the multiple linear model of hypothesis H2a (see Table 2), the dependent variable was the learning gain (or normalized change), and the independent variable was the three conditions of SE prompts (content-specific, guided, and generic). The hypothesis H2a would be confirmed if the coefficient of the guided condition is significantly greater than 0 (baseline) and the coefficient of the generic condition is significantly less than 0 (baseline) after controlling the individual difference variables. For the linear model of hypothesis H2b (see Table 2), the dependent variable is the proportion of high-quality self-explanations, and the independent variable is the three conditions. The hypothesis H2b would be confirmed if the coefficient of the guided condition is significantly greater than 0 (baseline) and the coefficient of the generic condition is significantly less than 0 (baseline) after controlling the individual difference variables. Then, the function *emmeans* in R was used to conduct a post-hoc analysis with Bonferroni correction.

Fleiss' κ was computed by the function *kappam.fleiss* in *irr* package in r to determine if there was an agreement between three raters' judgment on whether the self-explanations of a solution were considered as high quality (Fleiss, 1971). There was moderate agreement between the three raters, $\kappa = .513$ ($z = 56.6$, $p < 0.001$). After the discrepancy between the raters was resolved, the agreement between the final ratings and the ratings of rater 1 was $\kappa = .763$ ($z = 48.8$, $p < 0.001$), the agreement between the final ratings and the ratings of rater 2 was $\kappa = .762$ ($z = 48.7$, $p < 0.001$), and the agreement between the final ratings and the ratings of rater 2 was $\kappa = .737$ ($z = 47.0$, $p < 0.001$). The results indicated there was sufficient agreement among the three raters on the final ratings.

Table 2. *Multiple Linear Models to Test Hypotheses H2a/b*

	DV	**IV**	**Control Variables**
H2a	learning gains	3 SEP	age, gender, educational levels, and self-reported knowledge level on probability
H2b	proportion of high-quality SEs	3 SEP	

Note. SEP refers to the self-explanation prompting conditions. DV is the dependent variable. IV refers to the independent variables.

The multiple linear models with the same control variables were also used to test Hypotheses H3a and H3b (see Table 3). For hypothesis H3a, the interaction effects between prior knowledge and the three types of SE prompts (content-specific, generic, and guided) on the learning gains were investigated. The SE prompts conditions, prior knowledge as well as their interaction term were added to the multiple linear model as independent variables, while the learning gains (normalized change) of learners was the dependent variable. The content-specific condition was set as the baseline for three conditions in the model. Thus, the baseline for the interaction term was prior knowledge × content-specific. Then, the function *emmeans* in R would be used to conduct a post-hoc analysis with Bonferroni correction if the interaction effects were significant. The hypothesis H3a would be confirmed if the following three requirements are met:

1) The coefficient of the *guided* condition is greater than the coefficient of the *generic* condition.

2) There are no differences between the coefficients of the interaction terms *prior knowledge* × *guided* and *prior knowledge* × *generic or the coefficient*.

3) The coefficients of the interaction terms *prior knowledge* × *guided* and *prior knowledge* × *generic* are significantly greater than 0 (baseline).

For hypothesis H3b, the interaction effects between prior knowledge and the three types of SE prompts (Content-specific, Generic, and Guided) on the learning gains were investigated. Both the SE prompts conditions and learning ability as well as their interaction term were added into the model as independent variables and learning gains (or normalized change) of learners as dependent variable. The Content-specific condition was set as the baseline in the model. Thus, the baseline for the interaction term was *prior knowledge* × *content-specific*. Then, the function *emmeans* in R would be used to conduct a post-hoc analysis with Bonferroni correction. The hypothesis H3a would be confirmed if the following two requirements are met:

1) The coefficients of the interaction terms learning ability × generic and *learning ability* × *guided* are significantly greater than 0 (baseline).

2) The coefficient of the interaction term learning ability × generic is significantly greater than the coefficient of the learning ability × guided.

Table 3. *Multiple Linear Models to Test Hypotheses H3a/b*

	DV	IV	Control Variables
H3a	Learning gains	3 SEP, Prior Knowledge, & their interaction	age, gender, educational levels, and self-reported knowledge level on probability
H3b	Learning gains	3 SEP, Learning Ability, & their interaction	

Note. SEP refers to the self-explanation prompting conditions. DV is the dependent variable. IV refers to the independent variables.

A multiple linear regression model was used to test Hypothesis H4. The model used the number of high-quality self-explanations as well as other covariates (age, gender, educational levels, and the self-reported knowledge level on probability) to predict the learners' learning gains (normalized change) on far transfer tests (see Table 4). The hypothesis H4 would be confirmed if the coefficient of the number of high-quality self-explanations is significantly larger than 0.

Table 4. *Multiple Linear Regression to Test Hypotheses H4*

	DV	IV	Covariates
H4	learning gains on far transfer problems	the proportion of high-quality SEs	age, gender, educational levels, and the self-reported knowledge level on probability

Note. DV is the dependent variable. IV refers to the independent variables.

Results and Discussion

Descriptive Statistics

Individual Differences

The descriptive statistics of the individual difference variables used in the data analyses are reported across the three self-explanation prompt conditions. First, the learners' demographic data, self-reported knowledge level on probability, and the overall time they spent on the study were compared between the three conditions. learners' average overall time spent on the experiment, age, self-reported knowledge level on probability were shown in Table 5.

Table 5. *Learners' Age, Self-reported Knowledge Level on Probability and Overall Time on the Study*

	Generic (N=36)	**Guided (N=48)**	**Content-specific (N=45)**
Variable	**M (SD)**	**M (SD)**	**M (SD)**
Time (hrs.)	13.08 (29.90)	3.59 (1.49)	5.01 (4.97)
Age	35.14 (10.01)	33.27 (7.62)	33.24 (8.94)
Knowledge Level	3.06 (1.19)	3.31 (1.24)	3.31 (1.06)

On average, learners spent 6.84 (*SD* = 16.0) hours on the entire study. As shown in Table 5, the learners in the generic condition spent more time on the study than those in the guided condition and Content-specific condition. The learners in the guided condition spent the least time on the study. By checking the overall time that learners spent on the study, we found that there were 3 outliers greater than the value at 3 standard deviations above the mean. After removing the outliers, we conducted an ANOVA to compare the difference of the overall time on the study between the three self-explanation conditions at $p < 0.05$ level. There were no significant differences in overall time among the three conditions [$F(2, 123) = 2.26$, $p = 0.109$].

The learners of the study had an average age of 33.8 years old. As shown in Table 5, their age did not vary across the three self-explanation conditions. A one-way ANOVA on the age of the three conditions confirmed this observation. There was no significant difference in the age of the learners among the three conditions [$F(2, 126) = 0.593$, $p = 0.554$].

Table 6. *The Proportions of learners in Different Categories of Gender, Education Level, and Self-reported Knowledge Level on Probability across 3 Conditions*

Variable	Level	Generic (N=36)	Guided (N=48)	Content-specific (N=45)
Gender	Female	0.39	0.29	0.156
	Male	0.61	0.71	0.844
Education Level	High school degree or equivalent	0.03	0.10	0.000
	Bachelor's degree (e.g., BA, BS)	0.64	0.60	0.733
	Master's degree (e.g., MS, MA, MEd)	0.33	0.21	0.178
	Doctorate (e.g., PhD, EdD)	0.00	0.06	0.044
	Other	0.00	0.02	0.044
Knowledge Level	Not at all	0.08	0.08	0.022
	A little	0.28	0.19	0.244
	A moderate amount	0.28	0.27	0.267
	A lot	0.22	0.250	0.333
	Professional	0.14	0.21	0.133

From the *Learners* section, we know that the proportion of females was 0.27, and the rest were males. As shown in Table 6, among the 36 learners in the generic condition, the proportions of females and males were 0.39 and 0.61, respectively. Among the 48 learners in the guided condition, the proportions of females and males were 0.29 and 0.71, respectively. Among the 45 learners in the content-specific condition, the proportions of females and males were 0.16 and 0.84, respectively. A chi-square test of independence showed that there was no significant association between gender and three conditions, χ^2 (2, 129) = 5.67, p = 0.059.

The distribution of the learners' education levels across the three conditions was illustrated in Table 6. In the generic condition, the proportion of learners who had a high school degree or equivalent was 0.03, the proportion of learners who had a bachelor's degree (e.g., BA, BS) was 0.64, and the proportion of learners who had a master's degree (e.g., MS, MA, MEd) was 0.33. In the guided condition, the proportion of learners who had a high school degree or equivalent was 0.10, the proportion of learners who had a bachelor's degree (e.g., BA, BS) was 0.60, the proportion of learners who had a master's degree (e.g., MS, MA, MEd) was 0.21, the proportion of learners who had a doctorate degree (e.g., PhD, EdD) was 0.06, and the proportion of learners who fell in the "Other" category was 0.02. In the content-specific condition, the proportion of learners who had a bachelor's degree (e.g., BA, BS) was 0.73, the proportion of learners who had a master's degree (e.g., MS, MA, MEd) was 0.18, the proportion of learners who had a doctorate degree (e.g., PhD, EdD) was 0.04, and the proportion of learners who fell in the "Other" category was 0.04. A chi-square test of independence showed that there was no significant association between education levels and three conditions, χ^2 (8, 129) = 12.54, p = 0.129.

The self-reported knowledge levels on probability of the learners in the three conditions were both illustrated in Tables 5 and 6. As shown in Table 6, the proportion of learners who had no knowledge about probability at all was 0.08 in the generic condition, the proportion of learners who had "A little" knowledge about probability was 0.28, the proportion of learners who had "A moderate amount" knowledge about probability was 0.28, the proportion of learners who had "A lot" of knowledge about probability was 0.22, and the proportion of learners who reported themselves as professional in probability was 0.14. In the guided condition, the proportion of learners who had no knowledge about probability at all was 0.08, the proportion of learners who had "A little" knowledge about probability was 0.19, the proportion of learners who had "A moderate amount" knowledge about probability was 0.27, the proportion of learners who had "A lot" of knowledge about probability was 0.25, and the proportion of learners who reported themselves as professional in probability was 0.21. In the content-specific condition, the proportion of learners who had no knowledge about probability at all was 0.02, the proportion of learners who had "A little" knowledge about probability was 0.24, the proportion of learners who had "A moderate amount" knowledge about probability was 0.27, the proportion of learners who had "A lot" of knowledge about probability was 0.33, and the proportion of learners who reported themselves as professional in probability was 0.13. When taking the self-reported knowledge level on probability as an ordinal variable, we can calculate the average knowledge levels of learners in the three conditions. From table 5, we can see that learners in the generic condition had an average knowledge level of 3.06 (SD = 1.19), learners in the guided condition had an average knowledge level of 3.31 (SD = 1.24), and learners in the content-specific condition had an average knowledge level of 3.31 (SD = 1.06). A one-way between subjects ANOVA was conducted to compare learners' knowledge levels on probability of the learners in three self-explanation conditions. The results showed that no significant differences were observed among the three conditions [$F(2, 126) = 0.625$, $p = 0.537$]. Therefore, we can claim that, on average, the learners of the three conditions had "a moderate amount" (3) of knowledge on probability.

In summary, the learners in the three conditions had no significant differences in age, gender ratio, education levels, self-reported knowledge level of probability, and time that they spent on the study.

Learners' Performance

The learners' average proportions of accuracies on pretests and posttests of the two topics (*properties of probability* and *methods of enumeration*), overall learning gains, learning gains on the far transfer tests, as well as the proportions of learners' high-quality self-explanations in the three conditions were presented in Table 7 and 8. These variables served either as independent variables or as dependent variables for testing the hypotheses. The performance variables that were not used in the data analysis were ignored.

Table 7. *Learners' Performance on Topic I (Properties of Probability)*

	Generic (N=36)	Guided (N=48)	Content-specific (N=45)
	Mean (SD)	Mean (SD)	Mean (SD)
Prior Knowledge (Pretest)	0.76 (0.16)	0.67 (0.20)	0.72 (0.20)
Learning Ability (Pretest)	0.63 (0.14)	0.62 (0.20)	0.65 (0.19)
Posttest	0.66 (0.16)	0.61 (0.19)	0.59 (0.21)
Learning Gain	0.13 (0.31)	0.05 (0.31)	-0.06 (0.30)
Far Transfer (Pretest)	0.49 (0.20)	0.55 (0.26)	0.57 (0.25)
Far Transfer (Posttest)	0.54 (0.27)	0.52 (0.25)	0.46 (0.24)
Far Transfer (Learning Gain)	-0.11 (0.31)	-0.14 (0.31)	-0.24 (0.30)
High-quality Self-explanation	0.40 (0.30)	0.36 (0.31)	0.33 (0.31)

Table 8. *Learners' Performance on Topic II (Methods of Enumeration)*

	Generic (N=36)	Guided (N=48)	Content-specific (N=45)
	Mean (SD)	Mean (SD)	Mean (SD)
Prior Knowledge (Pretest)	0.62 (0.21)	0.57 (0.21)	0.56 (0.20)
Learning Ability (Pretest)	0.60 (0.18)	0.55 (0.17)	0.51 (0.17)
Posttest	0.57 (0.21)	0.57 (0.19)	0.47 (0.18)
Learning Gain	-0.01 (0.33)	0.08 (0.32)	-0.05 (0.26)
Far Transfer (Pretest)	0.61 (0.23)	0.54 (0.21)	0.47 (0.21)
Far Transfer (Posttest)	0.53 (0.27)	0.51 (0.25)	0.44 (0.21)
Far Transfer (Learning Gain)	-0.19 (0.33)	-0.15 (0.28)	-0.14 (0.27)
High-quality Self-explanation	0.45 (0.34)	0.41 (0.33)	0.34 (0.27)

Prior Knowledge (Pretest) was defined as a learner's prior declarative knowledge. It was measured by the proportion of their accuracy on the conceptual questions in the pretest of a topic. Learning Ability (Pretest) was defined as the amount of declarative knowledge learners could recall combined with their problem-solving performance after learning a topic. It was measured by the proportion of a learner's accuracy on the pretest of a topic. Posttest was the proportion of a learner's accuracy on the posttest of a topic. Learning Gain was a learner's normalized change from the pretest to the posttest of a topic. Far Transfer (Pretest) was the proportion of a learner's accuracy on the far transfer pretest of a topic. Far Transfer (Pretest) was the proportions of a learner's accuracy on the far transfer posttest of a topic. Far Transfer (Learning Gain) was a learner's normalized change from the far transfer pretest to the far transfer posttest of a topic. High-quality Self-explanation was the proportion of the high-quality self-explanations a learner generated when learning a topic.

Hypothesis H1: Are All Three Types of Self-explanation Prompts Effective?

One-tailed single sample *t*-tests were conducted to examine whether the learning gains (normalized change) of the learners in the three conditions (Generic, Guided, and Content-specific) were significantly greater than 0. The results are shown in Table 9. From the table, we can tell that learners in the generic condition had significant learning gains on Topic I (*properties of probability*). Although learners in the guided condition had positive learning gains that were greater than 0, the difference was not significant. Learners in the content-specific condition even had loss in learning gains, but the loss was not different from 0. Meanwhile, learners in the guided condition had significant learning gains on Topic II (*methods of enumeration*). Learners in the generic condition and the content-specific condition seemed to have negative learning gains, but their negative learning gains were not significantly different from 0 at $\alpha = 0.05$ level as shown in Table 9.

Table 9. *One-tailed T-tests on Whether the Learning Gains are Greater Than 0*

	Properties of Probability				**Methods of Enumeration**			
SE Conditions	Mean (SD) (LG)	t	df	p-value	Mean (SD) (LG)	t	df	p-value
Generic	**0.13 (0.31)**	**2.47**	**35**	**0.009****	-0.01 (0.33)	-0.218	35	0.414
Guided	0.05 (0.31)	1.03	47	0.155	**0.08 (0.32)**	**1.72**	**47**	**0.046***
Content-specific	-0.06 (0.30)	-1.24	44	0.097	-0.05 (0.26)	-1.39	44	0.086

Note. LG denotes the learning gains that were the normalized change from pretests to posttests. df is the degree of freedom. "*" indicates $p < 0.05$, "**" indicates $p < 0.01$.

The results of t-tests suggested that hypothesis H1 was partially supported. That is, not all types of computer-supported SE prompts were effective in improving learning probability. For topic I (properties of probability), generic SE prompts helped learners improve their learning by 13% of the maximum possible gain. Guided SE prompts seemed to promote learning, but the learning gains (normalized change) were not noteworthy. The content-specific SE prompts even prevented learners from learning because learners suffered from a notable loss in learning at α = 0.1 level of significance. For topic 2 (*methods of enumeration*), guided SE prompts helped learners improve learning by 8% of the maximum possible gain. The other two types of SE prompts caused some loss in learning topic 2. However, the loss caused by generic SE prompts was neglectable. The negative learning gains (normalized change) produced by content-specific SE prompts were significant at the α = 0.1 level.

These results suggested that generic SE prompts and guided SE prompts could improve learning or at least did not hinder learning, but content-specific SE prompts were not effective in both topics of probability. The effectiveness of the three types of SE prompts in the current study may be reduced by the sampling bias and fatigue effects. The learners were recruited from a crowdsourcing platform, Amazon Mechanical Turk. They were all adults but not real learners. It is probable that most of the learners were interested in learning some knowledge and in the meantime made some money. We cannot exclude that some learners were not interested in learning but only wanted to complete the tasks and get compensation. As a result, the effectiveness of these SE prompts was attenuated. Besides, it took many hours (generic: 13.5 hours on average, guided: 3.59 hours on average, and content-specific: 5.01 hours on average) for learners to complete the experiment. Learners might feel fatigued when they were working on the posttests. This is another reason that the effectiveness of these SE prompts was reduced. As was mentioned in the *Introduction* section, the content-specific SE prompts may lower the likelihood of learners generating a series of inferences because they direct learners' attention to specific content (Aleven et al., 2006). Based on the observation of self-explanations elicited by content-specific SE prompts, the content-related prompts not only directed learners' attention but also limited their attention to generating a series of inferences. For example, one content-related prompt was "What enumeration method(s) do you think was (were) used to solve this problem? Then explain why the solution is correct to yourself." Many learners only answered, "Multiplication rule was used" or "Permutation was used". They did not further explain why such methods should be used when solving the problem. Therefore, content-specific SE prompts hindered learning.

Hypotheses H2a/b: Are Guided Self-explanation Prompts Superior?

For hypothesis H2a, the results of the multiple linear regression models of the two topics of the subject matter are shown in Table 10. For the model of topic I (properties of probability), the R^2 value of 0.107 revealed that the predictors explained 10.7% of the variance with $F(6, 122)$

$= 2.39$. $p < 0.05$. The results of the model for topic I revealed that the education level of learners ($\beta = -0.17$, $p < 0.1$) had a significant negative relationship with their learning gains at the 0.1 *p*-level. Meanwhile, learners in the generic condition had significantly higher learning gains (normalized change) than those in the content-specific condition, $\beta = 0.24$, $p < 0.05$. And learners in the guided condition had higher learning gains (normalized change) than those in the Content-specific condition, $\beta = 0.14$, $p = 0.172$, but the difference did not attain significance at $\alpha = 0.05$. The results of the post-hoc analysis showed that learners' learning gains were not significantly different between the generic condition and the guided condition, $t(121) = 1.21$, $p = 0.348$. For the model of topic 2 (methods of enumeration), the R^2 value of 0.052 revealed that the predictors explained 5.2% of the variance with $F(6, 121) = 1.10$, $p = 0.36$. The results of the model for topic 2 revealed that individual difference variables (age, gender, educational levels, and the self-reported knowledge level on probability) did not predict learners' learning gains. However, learners in the guided condition had significantly higher learning gains (normalized change) than those in the Content-specific condition, $\beta = 0.14$, $p = 0.032$. The learning gains (normalized change) of learners in Guided and Content-specific conditions were not significantly different, $\beta = 0.04$, $p = 0.540$. The post-hoc analysis results showed that learners' learning gains were not significantly different between the generic condition and the guided condition, $t(118) = -1.39$, $p = 0.253$.

Table 10. *Regression on Learning Gains to Compare Learning Gains Differences on Three Conditions of Self-explanation Prompts*

	Properties of Probability					**Methods of Enumeration**				
Predictor	***β***	**B**	***SE*(B)**	***t***	***p*-value**	***β***	**B**	***SE*(B)**	***t***	***p*-value**
(Intercept)	0.00	-0.06	0.21	-0.27	0.788	-0.15	-0.15	0.21	-0.73	0.469
Age	0.15	0.01	0.00	1.58	0.116	0.00	0.00	0.00	1.10	0.273
Gender (F)	-0.07	-0.05	0.06	-0.73	0.468	0.04	0.04	0.06	0.70	0.485
Education	-0.17	-0.07	0.04	-1.90	0.060.	-0.01	-0.01	0.04	-0.37	0.716
KL	0.12	0.03	0.03	1.30	0.196	-0.01	-0.01	0.03	-0.19	0.847
SE (Generic)	**0.24**	**0.16**	**0.07**	**2.36**	**0.020***	0.04	0.04	0.07	0.61	0.540
SE (Guided)	0.14	0.09	0.06	1.37	0.172	**0.14**	**0.14**	**0.06**	**2.16**	**0.032***
	$R^2 = 0.107$					$R^2 = 0.052$				
	$F(6, 122) = 2.39$, $p < 0.05$*					$F(6, 121) = 1.10$, $p = 0.360$				

Note. For Gender, the baseline is Male. Education is an ordinal variable. KL denotes the self-reported knowledge level of probability. SE denotes self-explanation conditions. Content-specific condition is the baseline. "." indicates $p < 0.1$. "*" indicates $p < 0.05$.

The results of multiple linear regression models of the two topics did not fully confirmed the hypothesis *H2a* that the effectiveness of the three types of SE prompts follow the order of guided > content-specific > generic, but suggested that generic and guided SE prompts seemed to be more effective in promoting learning than content-specific SE prompts. The results of the regression model of topic I (*properties of probability*) revealed that generic SE prompts were significantly more effective than content-specific and that guided SE prompts were almost more effective than content-specific SE prompts at significance level of $\alpha = 0.1$. The results of the regression model of topic II (*methods of enumeration*) revealed that guided SE prompts were significantly more effective than content-specific SE prompts. Even though generic SE prompts were not significantly different from the content-specific SE prompts, they still produced higher learning gains (normalized change) than the content-specific SE prompts.

Both the generic SE prompts and the guided SE prompts use generic questions to elicit self-explanations. The only difference was that guided prompts provided some guidance about the common errors that learners may commit during probability problem-solving. Therefore, our results support the claim by Chi (2000) that generic prompts are more effective than content-specific or content-related prompts in improving learning because they enable learners to tailor their self-explanations for revising their own incomplete or incorrect knowledge structure or mental model. Generic prompts increase the opportunity for learners to detect gaps in their own understanding, discover deficiencies in their mental models of the learning contents, or generate useful inferences (Chi, 2000; VanLehn et al., 1992). However, content-specific prompts may not benefit the learners who already understood these contents. Even worse, they may prevent learners from generating a series of inferences because they direct learners' attention to specific content (Aleven et al., 2006).

Combining the results of Hypothesis H1 and hypothesis H2a, we found another interesting finding, namely that the effectiveness of different types of SE prompts may interact with different topics of the subject matter. We presumed that the number of concepts in the two topics and the difficulty levels of the two topics may have caused such interaction. By further observing the data, we found that learners answered a higher proportion of questions correctly in the pretest (M = 0.63, SD = 0.19) and posttests (M = 0.62, SD = 0.19) of topic I (*properties of probability*) than in pretest (M = 0.55, SD = 0.18) and posttests (M = 0.54, SD = 0.20) of topic II (*methods of enumeration*). Two sample Welch t-tests were performed to compare the performance accuracies of learners on the pretests and posttests between the two topics, respectively. It was confirmed that learners had higher performance accuracies on the pretest of topic I than the pretest of topic II, $t(255.4) = 3.70$, $p < 0.001$. They also had higher performance accuracies on the posttest of topic I than the posttest of topic II, $t(256) = 3.37$, $p < 0.001$. These results led to the conclusion that topic I may be easier than topic II. Topic I had more concepts, including set theory, Venn diagram, definition of events, properties of Set operations, definition of probability, some theorems of probability, etc., than topic II which only contained four concepts, multiplication, permutation, and sampling with/without replacement. Therefore, it is possible that learners benefit more from generic prompts when they study topics that include more concepts but are easy, whereas they benefit more from guided prompts when the topics include fewer concepts but are difficult. However, this assumption needs to be confirmed by further studies.

For hypothesis H2b, the results of the multiple linear regression models of the two topics were shown in Table 11. For the model of topic I (properties of probability), the R^2 value of 0.141 revealed that the predictors explained 14.1% of the variance, $F(6, 122) = 3.33$. $p < 0.01$. The results of the model for topic I revealed that learners' age ($\beta = 0.23$, $p < 0.05$) had a significant positive association with the numbers of high-quality self-explanations they generated, and learners' self-reported knowledge levels on probability ($\beta = -0.33$, $p < 0.001$) had a significant negative association with the numbers of high-quality self-explanations they generated. However, learners in the generic ($\beta = 0.08$, $p = 0.410$) and the guided ($\beta = 0.07$, $p =$

0.455) conditions did not generate significantly more high-quality self-explanations than those in the Content-specific condition. The results of post-hoc analysis showed that the numbers of the high-quality self-explanations that learners generated were not different between the generic condition and the guided condition, $t(122) = 0.145$, $p = 0.989$. For the model of Topic II (methods of enumeration), the R^2 value of 0.250 revealed that the predictors explained 25.0% of the variance with $F(6, 122) = 6.78$, $p < 0.001$. The results of the model for topic 2 revealed that learners' age ($\beta = 0.25$, $p < 0.01$) had significant positive association with the numbers of high-quality self-explanations they generated, and learners' self-reported knowledge levels on probability ($\beta = -0.20$, $p < 0.05$) had significant negative association with the numbers of high-quality self-explanations they generated. However, learners in the generic ($\beta = 0.11$, $p = 0.225$) and guided ($\beta = 0.12$, $p = 0.176$) conditions did not generate significantly more high-quality self-explanations than those in the content-specific condition. The results of post-hoc analysis showed that the numbers of the high-quality self-explanations learners generated were not different between the generic condition and the guided condition, $t(122) = -0.028$, $p > 0.1$.

Table 11. *Regression on the Proportions of High-Quality Self-explanations to Compare Their Differences on Three Conditions of Self-explanation Prompts*

	Properties of Probability					Methods of Enumeration				
Predictor	β	B	$SE(B)$	t	p-value	β	B	$SE(B)$	t	p-value
(Intercept)	0.00	0.25	0.20	1.29	0.199	0.00	0.31	0.19	1.66	0.100
Age	0.23	0.01	0.003	2.50	0.014*	0.25	0.01	0.003	2.96	0.004**
Gender (F)	0.13	0.09	0.06	1.46	0.146	0.13	0.09	0.06	1.61	0.110
Education	-0.06	-0.03	0.04	-0.73	0.469	-0.03	-0.01	0.03	-0.41	0.682
KL	-0.20	-0.05	0.02	-2.23	0.028*	-0.33	-0.08	0.02	-3.92	0.000***
SE (Generic)	0.08	0.05	0.07	0.83	0.410	0.11	0.08	0.06	1.22	0.225
SE (Guided)	0.07	0.05	0.06	0.75	0.455	0.12	0.08	0.05	1.36	0.176
	$R^2 = 0.141$					$R^2 = 0.250$				
	$F(6, 122) = 3.33$, $p < 0.01$**					$F(6, 122) = 6.78$, $p < 0.001$***				

Note. For Gender, the baseline is Male. Education is an ordinal variable. KL denotes the self-reported knowledge level of probability. SE denotes self-explanation conditions. Content-specific condition is the baseline. "*" indicates $p < 0.05$. "**" indicates $p < 0.01$. "***" indicates $p < 0.001$.

Hypothesis H2b predicted that the proportions of the high-quality self-explanations elicited by different types of prompts followed the order of guided > content-specific > generic. This hypothesis was not supported by the results of the multiple linear regression models.

However, the results demonstrated highly consistent patterns of data across the two topics of subject matter. That is, the number of high-quality self-explanations generated by learners increased with their age but decreased with their self-reported knowledge level on probability. Although the numbers of high-quality self-explanations generated by learners in the three conditions were not statistically different, there was still a trend that learners with generic and guided prompts generated more high-quality self-explanations than those with content-specific prompts.

It is important to explore why learners' age and self-reported knowledge level were related to the number of high-quality self-explanations they generated. A number of explanations may be considered as alternatives. First, maybe older learners were higher on conscientiousness (Robinson et al., 2021) and would like to fulfill their responsibility, e.g., completing the tasks in the experiment with compensation. Second, older learners may have less pre-existing knowledge of probability and more flaws in their mental model of the two topics of probability, which could be supported by the evidence that a significant negative correlation was found between learners' age and their self-reported knowledge levels on probability (r = -0.33, p < 0.001). Thus, they needed to generate more high-quality self-explanations to fill their gaps both in their mental model and the learning content.

The negative association between learners' self-reported knowledge levels on probability and high-quality explanations may suggest that learners with higher self-reported knowledge levels on probability probably had fewer gaps in their knowledge and mental model of the two topics of probability and believed that they did not have to explain much about the learning content. On the other hand, learners with low self-reported knowledge on probability had more gaps in their knowledge and mental model of the two topics and needed to generate more inferences to fill these gaps.

Although not significant, the coefficients of the *generic* and *guided* conditions were all greater than 0 (see Table 11). These results suggested that the generic and guided prompts may elicit 5-8% more high-quality self-explanations than the content-specific prompts. Combining the results of Hypothesis H2a, the generic and guided prompts may be superior to content-specific prompts both in producing learning gains and eliciting high-quality self-explanations. The moderate correlations (Topic I: $r = 0.39$, $p < 0.001$; Topic II: $r = 0.46$, $p < 0.001$) between learners' learning gains and the high-quality self-explanations they generated on the two topics implied that high-quality self-explanations were positively associated with learning gains.

Hypotheses 3a/b: Are There Interaction Effects between Learners' Aptitudes and SE Prompts?

The results of the models for hypothesis H3a are shown in Table 12. For the model of topic I (*properties of probability*), the R^2 value of 0.124 revealed that the predictors explained 12.4% of the variance with $F(9, 119) = 1.85$, $p < 0.1$. The results of the model revealed that the main effects of prior knowledge (PK) ($\beta = -0.05$, $p = 0.736$) and SE prompts (*generic*: $\beta = 0.63$, $p = 0.156$; *guided*: $\beta = 0.25$, $p = 0.502$) on learning gains (normalized change), and their interaction effects (*prior knowledge × generic*: $\beta = -0.39$, $p = 0.387$; *prior knowledge × guided*: $\beta = -0.12$, $p = 0.744$) were not significant after controlling learners' age, gender, education levels, self-reported knowledge levels of probability. These individual difference variables were not associated with the learning gains of learners on topic I. For the model of topic II (*methods of enumeration*), the R^2 value of 0.056 revealed that the predictors explained 5.6% of the variance with $F(9, 118) = 0.79$, $p = 0.630$. The results of the model revealed that the main effects of prior knowledge (PK) (β = 0.08, p = 0.633) and SE prompts (*generic*: β = 0.11, p = 0.742; *guided*: β = 0.44, p = 0.167) on learning gains (normalized change), and their interaction effects (*prior knowledge* × *generic*: β = -0.06, p = 0.870; *prior knowledge* × *guided*: β = -0.24, p = 0.467) were also not significant after controlling learners' age, gender, education levels, and self-reported knowledge levels of probability. These individual difference variables were not associated with the learning gains of learners on topic II.

The hypothesis H3a that there are interaction effects between prior knowledge and self-explanation prompts was not supported by our data. However, some suggestive patterns emerged from our data. Although not significant, the results of the regression models suggested that the interaction effects between prior knowledge and SE prompts followed a similar pattern across the two topics of subject matter. That is, compared to learners in the content-specific condition, the learning gains (normalized change) of learners in generic and guided conditions decreased as their prior knowledge increased. Since no interaction effects existed, the learning gains of learners in different conditions were not influenced by their prior knowledge. When combined with the results from hypothesis H2a, the learning gains of learners in the generic and guided conditions were higher than those in the content-specific condition regardless of their prior knowledge. These patterns were just an unconfirmed trend observed from our data and need further investigation.

Table 12. *Regression on Learning Gains to See the Interaction between Three Conditions of Self-explanation Prompts and Prior Knowledge*

	Properties of Probability					Methods of Enumeration				
Predictor	β	B	SE(B)	t	*p*-value	β	B	*SE*(B)	*t*	*p*-value
(Intercept)	0.00	-0.06	0.25	-0.23	0.816	0.00	-0.23	0.23	-0.99	0.322
Age	0.16	0.01	0.00	1.68	0.097	0.10	0.00	0.00	1.01	0.317
Gender (F)	-0.06	-0.04	0.06	-0.64	0.521	0.08	0.05	0.07	0.78	0.436
Education	-0.14	-0.06	0.04	-1.59	0.116	-0.04	-0.02	0.04	-0.40	0.688
KL	0.12	0.03	0.03	1.33	0.186	-0.02	0.00	0.03	-0.19	0.848
PK	-0.05	-0.01	0.02	-0.34	0.736	0.08	0.02	0.04	0.48	0.633
SE (G)	0.63	0.44	0.31	1.43	0.156	0.11	0.07	0.22	0.33	0.742
SE (U)	0.25	0.16	0.24	0.67	0.502	0.44	0.28	0.20	1.39	0.167
PK:SE (G)	-0.39	-0.03	0.03	-0.87	0.387	-0.06	-0.01	0.06	-0.16	0.87
PK: SE (U)	-0.12	-0.01	0.03	-0.33	0.744	-0.24	-0.04	0.06	-0.73	0.467
	$R^2 = 0.124$					$R^2 = 0.056$				
	$F(9, 119) = 1.85.$					$F(9, 118) = 0.79$				

Note. KL denotes the self-reported knowledge levels of probability. PK denotes prior knowledge. SE denotes the self-explanation conditions. G denotes Generic condition. U denotes the Guided condition. PK:SE(G) and PK:SE(U) denotes the interaction terms between prior knowledge and self-explanation conditions. "." indicates $p < 0.1$.

The results of the models for hypothesis H3b are shown in Table 13. For the model of topic I (*properties of probability*), the R^2 value of 0.137 revealed that the predictors explained 13.7% of the variance with $F(9, 119) = 2.10$, $p < 0.05$. The results of the model revealed that the main effects of learning ability (LA) ($\beta = -0.10$, $p = 0.502$) and SE prompts (*generic*: $\beta = 0.53$, $p = 0.173$; *guided*: $\beta = 0.25$, $p = 0.468$) on learning gains (normalized change), and their interaction effects (*learning ability* × *generic*: $\beta = -0.30$, $p = 0.438$; *learning ability* × *guided*: $\beta = -0.12$, $p = 0.725$) were not significant after controlling learners' age, gender, education levels, self-reported knowledge levels of probability. These individual difference variables were not associated with the learning gains of learners on topic I. For the model of topic II (*methods of enumeration*), the R^2 value of 0.070 revealed that the predictors explained 7.0% of the variance with $F(9, 118) = 0.79$, $p = 0.630$. The results of the model revealed that the main effects of learning ability (LA) ($\beta = -0.16$, $p = 0.295$) and SE prompts (*generic*: $\beta = 0.05$, $p = 0.894$; *guided*: $\beta = 0.44$, $p = 0.167$) on learning gains (normalized change), and their interaction effects (*learning ability* × *generic*: $\beta = 0.06$, $p = 0.877$; *learning ability* × *guided*: $\beta = 0.05$, $p = 0.899$) were either not significant after controlling learners' age, gender, education levels, self-reported knowledge levels of probability. These individual difference variables were also not associated with the learning gains of learners on topic II.

The hypothesis H3b that there are interaction effects between learning ability and self-explanation prompts was not supported by the data of this study. However, the interaction effects showed slightly different patterns between the two topics of the subject matter. For topic I, the learning gains (normalized change) of learners in the generic and guided conditions decreased as their learning ability increased compared to that of learners in the content-specific condition. For topic II, the learning gains (normalized change) of learners in the generic and guided conditions had an increasing trend as their learning ability increased compared to that of learners in the content-specific condition. It is possible that interaction effects exist between learning ability, self-explanation prompts, and difficulty levels of topics. Since Topic II was more difficult than Topic I, low learning ability learners may benefit more from generic and guided SE prompts when they learn less difficult topics, whereas high learning ability learners may benefit more from generic and guided SE prompts when they learn more difficult topics and vice versa. Again, these patterns were just an unconfirmed trend observed from our data and need further investigation.

Table 13. *Regression on Learning Gains to See the Interaction between Three Conditions of Self-explanation Prompts and Learning Ability*

	Properties of Probability					Methods of Enumeration				
Predictor	β	B	SE(B)	t	p-value	β	B	SE(B)	t	p-value
(Intercept)	0.00	-0.09	0.22	-0.41	0.683	0.00	-0.08	0.25	-0.33	0.739
Age	0.17	0.01	0.00	1.78	0.077.	0.15	0.01	0.00	1.44	0.152
Gender (F)	-0.05	-0.03	0.06	-0.54	0.590	0.10	0.07	0.07	1.01	0.314
Education	-0.14	-0.06	0.04	-1.63	0.106	-0.03	-0.01	0.04	-0.34	0.738
KL	0.13	0.03	0.03	1.39	0.168	-0.01	0.00	0.03	-0.09	0.930
LA	-0.10	-0.01	0.01	-0.67	0.502	-0.16	0.25	0.01	-1.05	0.295
SE (G)	0.53	0.37	0.27	1.37	0.173	0.05	0.03	0.24	0.13	0.894
SE (U)	0.25	0.16	0.22	0.73	0.468	0.20	0.13	0.22	0.57	0.570
LA:SE (G)	-0.30	-0.01	0.02	-0.78	0.438	0.06	0.00	0.02	0.15	0.877
LA: SE (U)	-0.12	0.00	0.01	-0.35	0.725	0.05	0.00	0.02	0.13	0.899
	$R^2 = 0.137$					$R^2 = 0.070$				
	$F(9, 119) = 2.10$*					$F(9, 118) = 0.98$				

Note. KL denotes the self-reported knowledge levels of probability. LA denotes learning ability. SE denotes the self-explanation conditions. G denotes Generic condition. U denotes the Guided condition. LA:SE(G) and LA:SE(U) denotes the interaction terms between learning ability and self-explanation conditions. "." indicates $p < 0.05$. "*" indicates $p < 0.05$.

The results of H3a and H3b of the current study showed that the “Aptitude-Treatment Interaction” (see Snow, 1991) may not exist for self-explanation prompts because learners’ aptitudes (prior knowledge and learning ability) did not alter the effects of self-explanation prompts on learning in general. However, based on the results, two interesting patterns of the data were observed. The first pattern was that prior knowledge might have identical effects on instructional treatments, e.g., self-explanation prompts, regardless of the difficulty levels of the subject matter, because the results showed that learners with low prior knowledge seemed to always benefit more from generic and guided SE prompts compared to content-specific SE prompts, and high prior knowledge learners might benefit more from content-specific prompts. Since prior knowledge was defined as declarative knowledge that includes the facts, concepts, principles (De Jong & Ferguson-Hessler, 1996), learners with low prior knowledge apparently had more flaws or incompleteness in their mental model of the probability topics than those with high prior knowledge. Therefore, they needed self-explanation to fill these gaps in order to understand the learning content better. Generic and guided prompts elicit learners’ self-explanation using content-free questions which increase the opportunity for learners to detect gaps in their own understanding, discover deficiencies in their mental models of the learning content, or generate useful inferences to make sense of the learning content (Chi, 2000; VanLehn et al., 1992). As a result, learners with low prior knowledge gained learning from generic and guided prompts. However, content-specific prompts could mislead low prior knowledge learners’ attention to some specific content that might not be the only missing parts in their knowledge structure (Aleven et al., 2006). As a result, content-specific prompts stop them to generate new ideas and hinder them from learning. Learners with high prior knowledge had fewer gaps in their understanding of the learning content, so the generic and guided prompts could hardly benefit them a lot without further assistance.

The other pattern was that the effects of learning ability on instructional treatments, e.g., self-explanation prompts, may vary with the difficulty levels of the topics of the subject matter. When the topic (topic II) was difficult, high learning ability learners may benefit more from the guided and generic SE prompts compared to content-specific SE prompts, whereas high learning ability learners did not differentially benefit from the three conditions. The learning ability was defined as the amount of declarative knowledge a learner retains and comprehends after learning a topic within a particular window of time. Learning ability involves the memory process, understanding of the learning content, and the ability to transfer the knowledge to new context. Learners with low learning ability had low declarative knowledge about the topics and limited ability to transfer the knowledge they learned. It was possible that they benefited from self-explanation with generic and guided prompts when they learned easy topics, e.g., topic I. However, they might be totally lost and unable to explain the learning content when learning difficult topics, e.g., topic II. In contrast, learners with high learning ability did not need help when they learned easy topics, but they needed self-explanation to help them better make sense of the difficult learning content.

Hypothesis 4: Does High-quality Self-Explanations Support Far Transfer?

The results of hypothesis 4 are shown in Table 14. For the model of topic I (properties of probability), the R^2 value of 0.110 revealed that the predictors explained 11.0% of the variance, $F(5, 123) = 3.04$. $p < 0.05$. The results of the model for topic I revealed that the number of high-quality self-explanations of learners ($\beta = 0.28$, $p < 0.01$) could predict their learning gains (normalized change) on far transfer tests after controlling their age, gender, education level, and self-reported knowledge level on probability. For the model of topic II (methods of enumeration), the R^2 value of 0.093 revealed that the predictors explained 9.3% of the variance, $F(5, 123) = 2.52$. $p < 0.01$. The results of the model for topic II revealed that the number of high-quality self-explanations of learners ($\beta = 0.28$, $p < 0.01$) could also predict their learning gains (normalized change) on far transfer problems after controlling their age, gender, education level, and self-reported knowledge level on probability. In both models, the individual difference variables were all not associated with the learning gains on far transfer tests.

The hypothesis H4 that high-quality self-explanations predict far transfer of the problem-solving procedures was supported by the results of the models of both topics. The results imply that the high-quality self-explanations help learners better fill the gaps of the learning content and their mental model of the learning content. As a result, learners can gain a deep understanding of the learning content and can apply the knowledge they learned in new or unfamiliar settings. This is consistent with the finding of Chi and colleagues' study (Chi et al., 1989) that high-quality self-explanations generated by learners are positively related to knowledge transfer.

Table 14. *High-quality Self-explanations Predict Learning Gains on Far Transfer Problems*

	Properties of Probability					Methods of Enumeration				
Predictor	***β***	**B**	***SE*(B)**	***t***	***p*-value**	***β***	**B**	***SE*(B)**	***t***	***p*-value**
(Intercept)	0.00	-0.13	0.22	-0.58	0.565	0.00	-0.14	0.19	-0.75	0.455
Age	0.07	0.00	0.00	0.76	0.446	-0.06	0.00	0.00	-0.66	0.512
Gender (F)	-0.07	-0.05	0.06	-0.81	0.418	0.10	0.06	0.06	1.09	0.279
Education	-0.11	-0.05	0.04	-1.29	0.200	-0.10	-0.04	0.03	-1.18	0.240
KL	0.02	0.00	0.02	0.20	0.841	0.05	0.01	0.02	0.50	0.620
HQSE	0.28	0.02	0.01	3.02	0.003**	0.28	0.02	0.01	2.86	0.005**
	$R^2 = 0.110$					$R^2 = 0.093$				
	$F(5, 123) = 3.04$, p =0.013*					$F(5, 123) = 2.52$, p = 0.033*				

Note. For Gender, the baseline is Male. Education is taken as an ordinal variable. KL denotes the self-reported knowledge level of probability. HQSE denotes high-quality self-explanations. "*" indicates $p < 0.05$. "**" indicates $p < 0.01$.

General Discussion

The current study first investigated whether the three types of computer-supported self-explanation prompts (generic, guided, and content-specific) were effective in improving learning, then compared their effectiveness in producing learning gains and generating high-quality self-explanations. Afterward, the interaction effects between learners' aptitudes (prior knowledge and learning ability) and three types of self-explanation prompts were explored. In the end, whether high-quality self-explanations support far transfer of problem-solving procedures was examined.

The results of hypothesis H1 and H2a suggested that generic and guided prompts were more effective in improving learning than content-specific prompts and they also produced significant learning gains. Both the generic prompts and the guided prompts use generic questions to elicit self-explanations. The only difference was that guided prompts provided some guidance that is sensitive to the common errors that learners may commit during probability problem-solving. Our results support the claim by Chi (2000) that generic prompts are more effective than content-specific or content-related prompts in improving learning because they enable learners to tailor their self-explanations for revising their own incomplete or incorrect knowledge structure or mental model (Chi et al., 1989; Chi et al., 1994). Generic prompts increase the opportunity for learners to detect gaps in their own understanding, discover deficiencies in their mental models of the learning contents, or generate useful inferences (Chi, 2000; VanLehn et al., 1992). In contrast, content-specific prompts may be helpful for some learners to realize that they have gaps in their understanding and even get hints to fill these gaps (VanLehn et al., 1992). However, such prompts may not benefit the learners who already understood these contents. Even worse, they may prevent learners from generating a series of inferences because they direct learners' attention to specific content (Aleven et al., 2006). The ineffectiveness of the content-specific prompts in the current study might be caused by the fact that these prompts misled learners' attention to specific content. Therefore, they deprived learners' opportunity to detect gaps in their mental model, which prevented them from generating useful inferences.

The inconsistent results of the models for the two probability topics may imply that there are interaction effects between learning content and different types of self-explanation prompts. That is, learners should adopt generic prompts when learning easy topics with many concepts and adopt generic prompts with some guidance when learning difficult topics with few concepts.

The results of hypothesis H2b revealed that the forms of the self-explanation prompts cannot predict whether the learners can generate high-quality self-explanations or not. However, the descriptive statistics suggested (non-significantly) that generic and guided prompts might have small advantages over content-specific prompts in eliciting high-quality self-explanations. That is, the pattern suggested they elicited more high-quality self-explanations than content-specific prompts. This potential explains why learners in the generic and guided conditions had higher learning gains. The correlational analysis also suggested a positive association between the learning gains and high-quality self-explanations.

No significant interaction effects between learners' aptitudes and different types of self-explanation prompts were found from the results of hypotheses H3a and H3b. This result supports the conclusion that learners' aptitudes (prior knowledge and learning ability) are not sensitive to the effects of different types of self-explanation prompts. This is inconsistent with the common findings of the Aptitude-Treatment Interaction (ATI) (see Snow, 1991). However, two interesting suggestive findings deserve further attention from researchers. First, prior knowledge may have identical sensitivity to instructional treatments. The results showed that learners with low prior knowledge benefited more from generic and guided SE prompts compared to those with high prior knowledge as content-specific prompts were not effective in general. How can this be explained? Learners with low prior knowledge had more gaps in their mental model of learning content than those with high prior knowledge. So, they needed self-explanation to fill these gaps. Generic and guided prompts that elicit self-explanations using content-free questions increased their opportunities to detect the gaps in their understanding, discover deficiencies in their mental models of the learning contents, or generate useful inferences (see Chi, 2000; VanLehn et al., 1992). As a result, learners with low prior knowledge gained learning from generic and guided prompts. However, content-specific prompts could mislead low prior knowledge learners' attention to some specific content that might not be the only missing parts in their knowledge structure (see Aleven et al., 2006). As a result, content-specific prompts stop them to generate new ideas and hinder them from learning. Learners with high prior knowledge had fewer gaps in their understanding of the learning content, so the generic and guided prompts might not benefit them a lot without further assistance.

The other finding was that the interaction effects between learning ability and instructional treatments may vary with the difficulty levels of the subject matter. Specifically, with the fact that content-specific prompts were generally ineffective, learners with low learning ability benefited more from generic and guided prompts when they learn easy topics and less from these prompts when they learn difficult topics, whereas learners with high learning were just the opposite. This may be because learners with low learning ability can easily make sense of the easy topics with self-explanation. However, they can be totally lost and unable to explain the difficult learning content. In contrast, learners with high learning ability do not need help when they learn easy topics but need self-explanation to help them better make sense of the difficult learning content.

The results of hypothesis H4 supported the claim that high-quality self-explanations support far transfer of problem-solving procedures that learners learned from the worked examples. The results indicate that the high-quality self-explanations help learners better fill the gaps of the learning content and increase the accuracy of their mental model of the learning content. The expected result is that learners have a deep understanding of the learning content and apply the knowledge they learned in new or unfamiliar settings. This is consistent with the work of Chi et al. (1989).

To sum up, the self-explanation prompts may be useful but not powerful interventions to support learning. Some types of self-explanation prompts, e.g., content-specific prompts, may sometimes have negative effects on learning. Learners' knowledge, skills, and aptitudes may not be sensitive to the effects of different types of self-explanation prompts on learning. However, high-quality self-explanations matter in producing high learning gains no matter how they are elicited. Finally, the field needs to further explore suggestive findings such as the interaction effects among learning ability, self-explanation prompts, and difficulty levels of the subject matter.

Limitations and Future Directions

There were three major limitations in the current study. The first limitation was related to sampling bias of the learners. The learners of the current study were all adults (their ages range from 21 to 60 years old) recruited from a crowd-sourcing platform, Amazon Mechanical Turk. They were not learners who really needed to learn the probability topics adopted in the experiment. They tended to be adults who were interested in math or probability or who thought themselves to be good at math or probability. A small number of them were apparently not interested in the probability. They merely tried to complete the tasks in the experiment and get the compensation. Therefore, it was questionable how many learners were cognitively engaged in the tasks when they were working on the experiment. Sampling bias and disengagement of learners might have reduced the effects of the self-explanation on learning and made the interaction effects between learners' aptitudes and self-explanation prompts insignificant.

The second limitation is that too many tasks (two reading materials, 4 tests including 90 questions, 32 solutions for self-explaining, and 1 or 2 training sessions) in the experiment made the learners spend a great amount of time (6.84 hours on average) to complete. Such intense cognitive activities presumably are correlated with fatigue effects. Learners might cognitively engage in the tasks at the beginning of the experiment, but as time went on, they felt fatigue and could not fully engage their cognitive resources into the tasks. One evidence of fatigue effects is when learners have negative learning gains, that is, they did better on the pretests than on the posttests. The fatigue effects could also reduce the effects of the self-explanation on learning and rendered the interaction effects between learners' aptitudes and self-explanation prompts insignificant.

The third limitation is that the intervention of self-explanation on learning was not long enough. Acquiring conceptual knowledge and the ability of transferring the knowledge into new or unfamiliar settings is a long process. A few hours of learning were far from enough, which might be the reason that the average learning gains (normalized change) were small and occasionally negative.

There may be other limitations. For example, the training materials of how to generate high quality self-explanations and how to identify the common errors that learners may commit during probability problem-solving may be so short that the learners did not fully understand the materials. The questions used in content-specific prompts might mislead learners' attention and prevent them from generating high quality self-explanations.

Future research should first resolve the limitations of the current study. There is one way that can resolve all the major limitations. That is, learners should be sampled from real learners who are studying these probability topics in high schools or colleges. When the real learners are learning these probability topics, they do not intensively study them in several hours. Teachers always distribute the learning content of these topics into several days or weeks so that the learners will not get tired or experience fatigue effects. Meanwhile, the bona fide learners have the motivation to learn these probability topics. Researchers could potentially add a self-explanation session while the learners are doing their homework using computer-supported systems, e.g., AutoTutor (see Graesser, 2016).

The suggestive findings from the current study provide some new directions for future research on self-explanation. For example, the interaction effects between learners' aptitudes, different types of self-explanation prompts, and difficulty levels of learning content need further investigation. Such research could provide insights about individualized use of different types of self-explanation prompts and how they interact with different difficulty levels of learning content to learning scientists and designers of the computer-based learning systems. Many research questions can be asked following this vein. For example, should learners adopt generic prompts when learning easy topics with many concepts and adopt generic prompts with some guidance when learning difficult topics with few concepts? Do generic prompts benefit learners with low prior knowledge when they learn easy topics? Do guided prompts benefit learners with high learning ability when they learn difficult topics? Another branch of research can focus on what learning content learners should explain in order to get the most benefit from self-explanation. In essence, what are the best worked examples for a learner to explain?

Future research can also investigate the interaction effects between different types of self-explanation prompts and other characteristics of the learners, for example, cognitive style, personality traits, and so on. It would be interesting to see, for example, whether the personality trait of grit (see Duckworth et al., 2007) can moderate the effectiveness of self-explanations. The suggested directions of self-explanation research will deepen our understanding of the use of self-explanation prompts, enrich the theories of learning sciences, and provide theoretical support for prompting self-explanations in intelligent tutoring systems and other computer-supported learning environments.

Reference

Aizikovitsh, E., & Amit, M. (2008). Developing Critical Thinking in Probability Session. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepúlveda (Eds.), *Proceedings of the Joint Meeting of PME 32 and PME-NA XXX* (Vol. 2, pp. 9-16). Cinvestav-UMSNH.

Aleven, V., & Koedinger, K. R. (2000). The need for tutorial dialog to support self-explanation. In C. P. Rose & R. Freedman (Eds.), *Building Dialogue Systems for Tutorial Applications, 2000 AAAI Fall Symposium* (pp. 65-73). AAAI Press.

Aleven, V., Ogan, A., Popescu, O., Torrey, C., & Koedinger, K. (2004). Evaluating the effectiveness of a tutorial dialogue system for self-explanation. In J. C. Lester, R. M.

Vicari, & F. Paraguaçu (Eds.), *Proceedings of the 7th International conference on intelligent tutoring systems* (pp. 443-454). Springer.

Aleven, V., Pinkwart, N., Ashley, K., & Lynch, C. (2006). Supporting self-explanation of argument transcripts: Specific v. generic prompts. In V. Aleven, K. Ashley, C. Lynch, & N. Pinkwart (Eds.), *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains Held during the 8th International Conference on Intelligent Tutoring Systems* (pp. 47-55).

Aleven, V., Popescu, O., & Koedinger, K. R. (2001). Towards tutorial dialog to support self-explanation: Adding natural language understanding to a cognitive tutor. In J. D. Moore, C. L. Redfield, & W. L. Johnson (Eds.), *Proceedings of the 10th International Conference on Artificial Intelligence in Educationn* (pp. 246-255). Citeseer.

Aleven, V., Popescu, O., Ogan, A., & Koedinger, K. R. (2003). A formative classroom evaluation of a tutorial dialogue system that supports self-explanation. In V. Aleven (Ed.), *Supplementary Proceedings of 11th Int Conf on Artificial Intelligence in Education* (Vol. 6, pp. 345-355).

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer‐based cognitive tutor. *Cognitive science*, *26*(2), 147-179.

Ang, L. H., & Shahrill, M. (2014). Identifying learners' specific misconceptions in learning probability. *International Journal of Probability and Statistics*, *3*(2), 23-29.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational research*, *70*(2), 181-214.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of educational psychology*, *95*(4), 774.

Bereiter, C., & Scardamalia, M. (1985). Cognitive coping strategies and the problem of "inert knowledge". In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Research and Open Questions* (Vol. 2, pp. 65-80). Routledge.

Berthold, K., Eysink, T. H., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*(4), 345-363.

Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. *Computers in Human Behavior*, *27*(1), 69-75.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and instruction*, *13*(2), 221-252.

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, *30*, 703-725.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, *13*(6), 4-16.

Booth, J. L., Cooper, L. A., Donovan, M. S., Huyghe, A., Koedinger, K. R., & Paré-Blagoev, E. J. (2015). Design-based research within the constraints of practice: AlgebraByExample. *Journal of Education for Learners Placed at Risk*, *20*(1-2), 79-100.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve learner learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24-34.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *Hou people learn II: Brain, mind, experience, and school*. Washington: National Academy Press.

Broers, N. J., & Imbos, T. (2005). Charting and manipulating propositions as methods to promote self-explanation in the study of statistics. *Learning and Instruction*, *15*(6), 517-538.

Canfield, W. (2001). ALEKS: A Web-based intelligent tutoring system. *Mathematics and Computer Education*, *35*(2), 152.

Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction*, *16*(2), 139-153.

Chi, M. T. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in instructional psychology*, *5*, 161-238.

Chi, M. T. (2009). Active - constructive - interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, *1*(1), 73-105.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How learners study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145-182.

Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, *18*(3), 439-477.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive science*, *5*(2), 121-152.

Chiu, J. L., & Chi, M. T. (2014). Supporting self-explanation in the classroom. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 91-103). Society for the Teaching of Psychology.

Chou, C.-Y., & Liang, H.-T. (2009). Content-free computer supports for self-explaining: Modifiable typing interface and prompting. *Journal of Educational Technology & Society, 12*(1), 121-133.

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *The Journal of Philosophy for Children*, *8*(1), 2-10.

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*, 389-415.

Cote, N. (1994). *Overcoming the Inert Knowledge Problem in Learning from Expository Text* [Unpulished Manuscript]. Vanderbilt University.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Davis, E. A. (2003). Prompting middle school science learners for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, *12*(1), 91-142.

de Bruin, A. B., Rikers, R. M., & Schmidt, H. G. (2007). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary educational psychology*, *32*(2), 188-205.

De Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105-113.

De Jong, T., & Lazonder, A. W. (2014). The Guided Discovery Learning Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed., pp. 371-385). Cambridge University Press.

De Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2011). Improved effectiveness of cueing by self‐explanations when learning from a complex animation. *Applied Cognitive Psychology*, *25*(2), 183-194.

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of experimental child psychology*, *113*(4), 552-568.

Duckworth, A. L., Peterson, C., Matthews, M. D., Kelly, D. R. J. J. o. p., & psychology, s. (2007). Grit: perseverance and passion for long-term goals. *92*(6), 1087.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, *22*(3), 206-214.

Eiriksdottir, E., & Catrambone, R. (2011). Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human factors*, *53*(6), 749-770.

Eysink, T. H., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal*, *46*(4), 1107-1149.

Falk, R., & Konold, C. (1992). The psychology of learning probability. *J Statistics for the twenty-first century*, 151-164.

Fonseca, B. A., & Chi, M. T. (2011). Instruction based on self-explanation. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (2 ed., pp. 296-321). Routledge.

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*(1), 47-61.

Gholson, B., & Craig, S. D. (2006). Promoting constructive activities that support vicarious learning during computer-based instruction. *Educational Psychology Review*, *18*(2), 119-139.

Graesser, A. C. (2008). *25 Learning Principles to Guide Pedagogy and the Design of Learning Environments*. http://www.bgsu.edu/downloads/provost/file47947.pdf

Graesser, A. C. (2016). Conversations with AutoTutor help learners learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124-132.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*(4), 612-618.

Graesser, A. C., Hu, X., & Sottilare, R. (2018). Intelligent tutoring systems. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 246-255). Routledge.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 180-192.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one‐to‐one tutoring. *Applied Cognitive Psychology*, *9*(6), 495-522.

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*(1), 93-103.

Haskell, R. E. (2001). *Transfer of learning: Cognition, instruction, and reasoning*. Academic Press. https://doi.org/10.1016/B978-012330595-4/50003-2

Hausmann, R. G., & Chi, M. H. (2002). Can a computer interface support self-explaining. *Cognitive Technology*, *7*(1), 4-14.

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, *18*(1), 54-65.

Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2010). *Probability and statistical inference* (9 ed.). Pearson.

Hu, X., Xu, Y. J., Hall, C., Walker, K., & Okwumabua, T. (2013). A potential technological solution for reducing the achievement gap between White and Black learners. In J. C. Falmagne, D. Albert, C. Doble, D. Eppstein, & X. Hu (Eds.), *Knowledge spaces* (pp. 79-91). Springer. https://doi.org/https://doi.org/10.1007/978-3-642-35329-1_5

Huang, X., Craig, S. D., Xie, J., Graesser, A., & Hu, X. (2016). Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program. *Learning and Individual Differences*, *47*, 258-265.

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior* *26*(6), 1246-1252.

Kalyuga, S. (2009). Instructional designs for the development of transferable knowledge and skills: A cognitive load perspective. *Computers in Human Behavior*, *25*(2), 332-338.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1-19.

Kevan, J. M., & Ryan, P. R. (2016). Experience API: Flexible, decentralized and activity-centric data collection. *Technology, knowledge, and Learning*, *21*(1), 143-149.

Khazanov, L., & Prado, L. (2009). Instructors' perspectives on learners' mistaken beliefs about probability in an elementary college statistics course. *ALM International Journal*, *5*(1), 23-35.

Khazanov, L., & Prado, L. (2010). Correcting Learners' Misconceptions about Probability in an Introductory College Statistics Course. *Adults Learning Mathematics*, *5*(1), 23-35.

King, A. (1991). Improving lecture comprehension: Effects of a metacognitive strategy. *Applied Cognitive Psychology*, *5*(4), 331-346.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, *85*(5), 363.

Kintsch, W., & Vipond, D. (2014). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory* (pp. 329-365). Routledge.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86.

Koedinger, K. R., Anderson, J. R., Hadley, W., & Mark, M. A. (1997). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education (IJAIED)*, *8*, 30-43. https://doi.org/10.1184/R1/6470153.v1

Koedinger, K. R., & Corbett, A. (2006). Cognitive Tutors: Technology Bringing Learning Sciences to the Classroom. In *The Cambridge handbook of: The learning sciences.* (pp. 61-77). Cambridge University Press.

Kramarski, B., Weiss, I., & Sharon, S. (2013). Generic versus context-specific prompts for supporting self-regulation in mathematical problem solving among learners with low or high prior knowledge. *Journal of Cognitive Education and Psychology*, *12*(2), 197-214.

Kuhn, D., & Katz, J. (2009). Are self-explanations always beneficial? *Journal of experimental child psychology*, *103*(3), 386-394.

Kulik, J. A., & Fletcher, J. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational research*, *86*(1), 42-78.

Kurtz, K. J., & Honke, G. (2020). Sorting out the problem of inert knowledge: Category construction to promote spontaneous transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 803.

Kwon, K., Kumalasari, C. D., & Howland, J. L. (2011). Self-Explanation Prompts on Problem-Solving Performance in an Interactive Learning Environment. *Journal of Interactive Online Learning*, *10*(2), 96-112.

Kyllonen, P. C., & Lajoie, S. P. (2003). Reassessing aptitude: Introduction to a special issue in honor of Richard E. Snow. *Educational Psychologist*, *38*(2), 79-83.

Leppink, J., Broers, N. J., Imbos, T., van der Vleuten, C. P., & Berger, M. P. (2012). Self-explanation in the domain of statistics: an expertise reversal effect. *Higher Education*, *63*(6), 771-785.

Margulieux, L. E., & Catrambone, R. (2019). Finding the best types of guidance for constructing self-explanations of subgoals in programming. *Journal of the Learning Sciences*, *28*(1), 108-151.

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, *75*(1), 87-91.

McEldoon, K. L., Durkin, K. L., & Rittle‐Johnson, B. (2013). Is self‐explanation worth the time? A comparison to additional practice. *British Journal of Educational Psychology*, *83*(4), 615-632.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse processes*, *38*(1), 1-30.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, *14*(1), 1-43.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 222-233.

McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, technologies* (pp. 397-421). Lawrence Erlbaum Associates.

Mitsea, E., & Drigas, A. (2019). A Journey into the Metacognitive Learning Strategies. *International Journal of Online and Biomedical Engineering (iJOE)*, *15*(14), 4-20.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and instruction*, *16*(2), 173-199.

NCTM. (2000). *Principles and Standards for School Mathematics.* NCTM.

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427-469.

Nye, B. D., Rahman, M. F., Yang, M., Hays, P., Cai, Z., Graesser, A., & Hu, X. (2014). *A tutoring page markup suite for integrating Shareable Knowledge Objects (SKO) with HTML*. Intelligent Tutoring Systems (ITS) 2014 Workshop on Authoring Tools,

O'Connell, A. A. (1999). Understanding the nature of errors in probability problem-solving. *Educational Research and Evaluation*, *5*(1), 1-21.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Learner Learning.* National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. http://ncer.ed.gov/

Pirolli, P., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and instruction*, *12*(3), 235-275.

Recker, M. M., & Pirolli, P. (1995). Modeling individual differences in learners' learning strategies. *The Journal of the Learning Sciences*, *4*(1), 1-38.

Renkl, A. (1997). Learning from worked‐out examples: A study on individual differences. *Cognitive science*, *21*(1), 1-29.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529-556.

Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229-245). Cambridge University Press.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15-22.

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*(2), 115-121.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary educational psychology*, *23*(1), 90-108.

Ritter, S., Anderson, J. R., Koedinger, K. R., Corbett, A. J. P. b., & review. (2007). Cognitive Tutor: Applied research in mathematics education. *14*(2), 249-255.

Rittle-Johnson, B., & Loehr, A. M. (2017). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic bulletin & review*, *24*(5), 1501-1510.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, *49*(4), 599-611.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology*, *93*(2), 346.

Rittle‐Johnson, B. (2006). Promoting transfer: Effects of self‐explanation and direct instruction. *Child development*, *77*(1), 1-15.

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *Journal of Education*, *196*(2), 1-38.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426-445.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of educational psychology*, *99*(2), 285.

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31-58). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511489709

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of consulting clinical psychology*
59(2), 205.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, *36*(5), 404-411.

Star, J. R. (2007). Foregrounding Procedural Knowledge. *Journal for Research in Mathematics Education*, *38*(2), 132-135. https://doi.org/10.2307/30034953

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (1998). Instructional effects in complex learning: Do objective and subjective learning outcomes converge? *Learning and Instruction*, *8*(2), 117-129.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college learners' academic learning. *Journal of educational psychology*, *106*(2), 331.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123-138.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and instruction*, *2*(1), 59-89.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary educational psychology*, *36*(3), 212-218.

VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of CASCADE. *The Journal of the Learning Sciences*, *8*(1), 71-125.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.

VanLehn, K., Jones, R. M., & Chi, M. T. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, *2*(1), 1-59.

Whitehead, A. N. (1955). *The aims of education and other essays*. Williams and Norgate.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, *142*(4), 1006.

Wolfe, M. B., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and instruction*, *23*(4), 467-502.

Woltz, D. J. (2018). Implicit cognitive processes as aptitudes for learning. In *Educational Psychologist* (pp. 95-104). Routledge.

Wong, R. M., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on learners' problem solving in high-school mathematics. *Learning and Instruction*, *12*(2), 233-262.

Wylie, R., & Chi, M. T. H. (2014). The Self-Explanation Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed., pp. 413-432). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139547369.021

Appendices

Appendix I Learning Materials

1. Slides for Properties of Probability

https://docs.google.com/presentation/d/e/2PACX-1vRWD5_neiLo52N-cSd5cvRKSfnswbH68okztXJWVkSrkkJtYc97-PJNICzYWdrtBw7n8G7xpnIa3Fhs/pub?start=false&loop=false&delayms=3000

2. Slides for Methods of Enumeration

https://docs.google.com/presentation/d/e/2PACX-1vQKslM-Z6Y76ihzLq-5RmGs62W09a1Gm72fozvx-gATrJz3Wcu9k8TxEw1U1W18AJGPbuLoAzSGr31d/pub?start=false&loop=false&delayms=3000

3. Self-explanations Session for Properties of Probability

Generic Prompts[2]: https://app.skoonline.org/GHS/SKO/Framed.html?guid=e6db3351-c002-48b5-9c0a-f9f2b3a64653

Content-specific Prompts: https://app.skoonline.org/GHS/SKO/Framed.html?guid=bf61aeac-f07a-43d0-b383-6e961d12e8aa

4. Self-explanation Session for Methods of Enumeration

Generic Prompts[2]: https://app.skoonline.org/GHS/SKO/Framed.html?guid=1f2c588a-0ebd-42e3-9bea-943233c74967

Content-specific Prompts: https://app.skoonline.org/GHS/SKO/Framed.html?guid=f71578eb-58b9-4931-8f6b-49b04436b737

5. Tutorial of How to Use AutoTutor

https://app.skoonline.org/GHS/SKO/Framed.html?guid=a38221bf-6d22-412d-bc15-96c5ad6def53

6. Training Materials of Errors

https://app.skoonline.org/GHS/SKO/Framed.html?guid=70d18af0-0802-4a48-ac61-605f3bbda5fe

[2] The generic condition and the guided condition used self-explanation sessions with generic prompts.

Appendix 1I Demographic Survey

Q1 How old are you?

____

Q2 What is your gender?

- ○ Female (1)
- ○ Male (2)

Q3 What is the highest degree or level of school you have completed?

- ○ Less than a high school diploma (1)
- ○ High school degree or equivalent (2)
- ○ Bachelor's degree (e.g., BA, BS) (3)
- ○ Master's degree (e.g., MS, MA, MEd) (4)
- ○ Doctorate (e.g., PhD, EdD) (5)
- ○ Other (Please specify) (6) ____

Q4 How much knowledge do you have about Probability?

- ○ None at all (1)
- ○ A little (2)
- ○ A moderate amount (3)
- ○ A lot (4)
- ○ Professional (5)

Appendix III Tests

Test I for Properties of Probability

Q1.1 ∅ denotes the _____.

- ◯ empty set
- ◯ full set
- ◯ subset
- ◯ super set

Q2.1 A ⊂ B means A is a ________ of B.

- ◯ empty set
- ◯ full set
- ◯ subset
- ◯ super set

Q3.1 A ∪ B means _________.

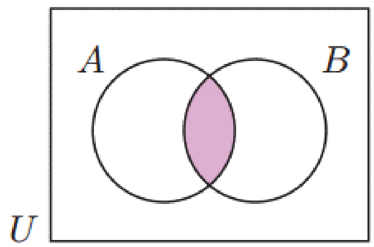
- ◯ A is a subset of B
- ◯ B is subset of A
- ◯ A intersect B
- ◯ A union B

Q4.1 A′ is the __________ of A.

- ◯ subset
- ◯ complement
- ◯ space
- ◯ probability

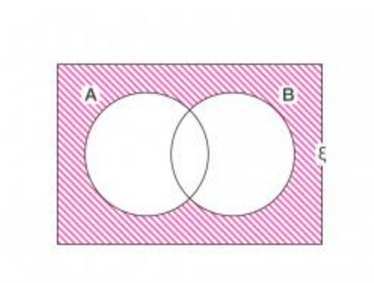
Q5.1 What statement does the shaded region represent?

- ◯ A ∪ B
- ◯ A'
- ◯ A ∩ B
- ◯ B'



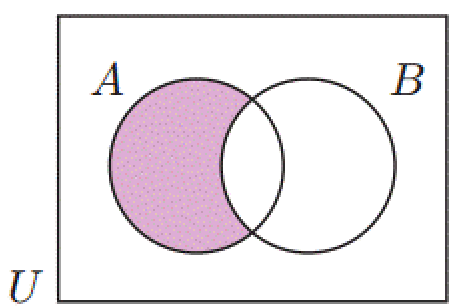
Q5.3 Which region is represented by the diagram?

- ◯ A' ∩ B
- ◯ A' ∩ B'
- ◯ A' ∪ B'
- ◯ A ∪ B'



Q5.5 What statement does the shaded region represent?

- ◯ A ∩ B'
- ◯ B'
- ◯ A'
- ◯ A ∪ B'



Q6.1 A1, A2, ... Ak are __________ events mean that Ai ∩ Aj = ∅; that is, A1, A2, ... Ak are disjoint sets.

- ◯ mutually exclusive
- ◯ exhaustive
- ◯ mutually exclusive and exhaustive
- ◯ exclusive

Q7.1 A ∪ (B ∩ C) = __________.

- ◯ (A ∩ B) ∪ (A ∩ C)
- ◯ (A ∪ B) ∩ (A ∪ C)
- ◯ (A ∪ B) ∩ (B ∪ C)
- ◯ (A ∩ B) ∪ (A ∪ C)

Q8.1 (A ∪ B)' = _________.

- ◯ A' ∪ B'
- ◯ A' ∩ B'
- ◯ A ∩ B
- ◯ A ∪ B

Q9.1 The probability of event A, denoted by P(A) is often called the _______ of event A occurring.

Q10.1 If events A and B are such that A ⊂ B, then ____________.

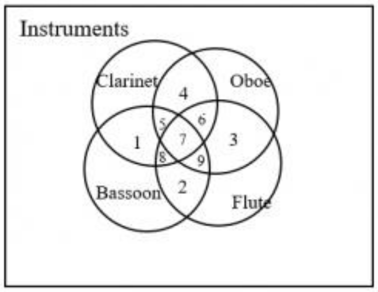
- ◯ P(A) ≥ P(B)
- ◯ P(A) ≤ P(B)
- ◯ P(A) > P(B)
- ◯ P(A) < P(B)

Q11.1 P(A ∪ B) = _______________.

- ◯ P(A) + P(B)
- ◯ P(A) + P(B) - P(A ∩ B)
- ◯ P(A) + P(A ∩ B)
- ◯ P(B) + P(A ∩ B)

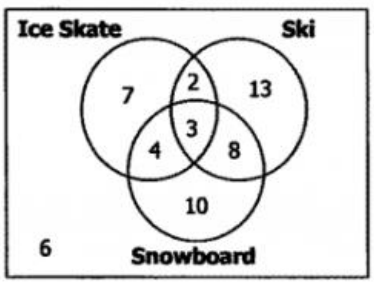
Q20.1 Identify each region of the Venn diagram that represents learners who play only the clarinet and oboe.

- ◯ 4
- ◯ 4+5+6
- ◯ 4+5+6+7
- ◯ 1+3+4+5+6+7+8+9

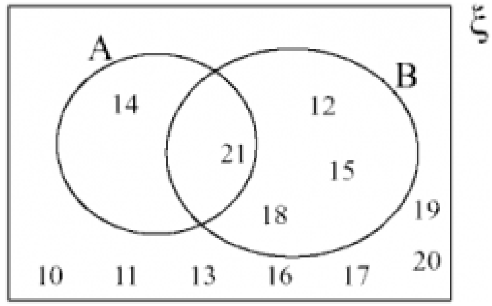


Q21.1 How many learners do not snowboard?

- ◯ 21
- ◯ 22
- ◯ 28
- ◯ 37



Q22.1 Which is the correct set notation for A U B?



- ◯ {21}
- ◯ {12, 14, 15, 18, 21}
- ◯ {10, 11, 13, 16, 17, 19, 20}
- ◯ {12, 14, 15, 18}

Q23.1 At a breakfast buffet, 93 people chose coffee and 47 people chose juice. 25 people chose both coffee and juice. If each person chose at least one of these beverages, how many people visited the buffet?

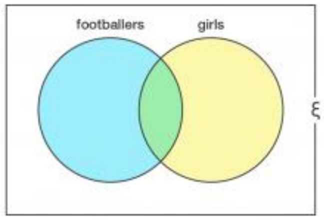
- ◯ 118
- ◯ 165
- ◯ 115
- ◯ 93

Q24.1 In a class of 30 learners, 19 are studying Chinese, 12 are studying Spanish and 7 are studying both Chinese and Spanish. How many learners are not taking any foreign languages?

- ◯ 6
- ◯ 12
- ◯ 24
- ◯ 0

Q25.1 In a class of 30 learners, 15 learners play football. 7 boys don't play football and 6 girls do play football. How many boys are there in the class?

- 9
- 16
- 14
- 23



Q30.1 Two events (A and B, [illegible] utually exclusive. The probability that neither occur is

- 0
- 0.4
- 0.04
- 0.6
- none of the preceding

Q31.1 A smoke-detector system consists of two parts A and B. If smoke occurs then the item A detects it with probability 0.95, the item B detects it with probability 0.98 whereas both of them detect it with probability 0.94. What is the probability that the smoke will not be detected?

- 0.01
- 0.99
- 0.04
- 0.96
- none of the preceding

Q32.1 The probability that a learner passes Statistics course is 2/3 and the probability that he passes both Statistics and mathematics course is 14/45. The probability that he passes at least one course is 4/5. what is the probability that he passes mathematics course?

- 2/15
- 4/9
- 18/135
- 112/135

Q34.1 Of a group of patients having injuries, 28% visit both a physical therapist and a chiropractor and 8% visit neither. Say that the probability of visiting a physical therapist exceeds the probability of visiting a chiropractor by 16%. What is the probability of a randomly selected person from this group visiting a physical therapist?

- 0.54
- 0.68
- 0.52
- 0.22

Q35.1 An insurance company looks at its auto insurance customers and finds that (a) all insure at least one car(b) 85% insure more than one car(c) 23% insure a sports car(d) 17% insure more than one car, including a sports car. Find the probability that a customer selected at random insures exactly one car and it is not a sports car.

- 0.06
- 0.09
- 0.68
- 0.91

Q36.1 During a visit to a primary care physician's office, the probability of having neither lab work nor referral to a specialist is 0.21. Of those coming to that office, the probability of having lab work is 0.41 and the probability of having a referral is 0.53. What is the probability of having both lab work and a referral?

- ○ 0.79
- ○ 0.26
- ○ 0.15
- ○ 0.38

Test II for Properties of Probability

Q1.2 What symbol denotes an empty or null set?

- ○ ∅
- ○ ⊂
- ○ ∪
- ○ S

Q2.2 A ⊃ B means __________.

- ○ A is a subset of B
- ○ B is a subset of A
- ○ A intersect B
- ○ A union B

Q3.2 A ∩ B means __________ .

○ A is a subset of B

○ B is a subset of A

○ A intersect B

○ A union B

Q4.2 The complement of A is _________.

○ A'

○ S

○ B

○ ∅

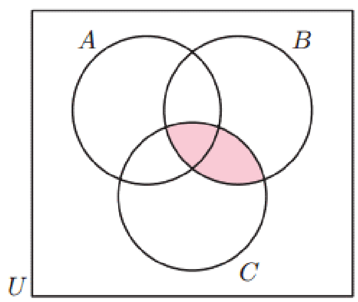
Q5.2 What statement does the shaded region represent?

○ A ∩ B ∩ C

○ A ∪ B ∩ C

○ B ∩ C

○ A ∪ C



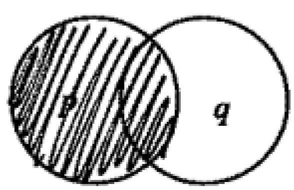
Q5.4 What does the shaded portion of the Venn diagram represent?

○ p'

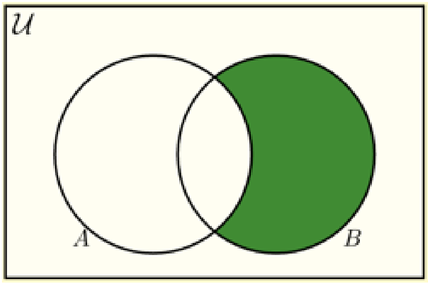
○ p

○ p ∩ q

○ p ∪ q



Q5.6 What statement does the shaded region represent?



- ◯ A' ∩ B
- ◯ A' ∪ B
- ◯ A'
- ◯ B'

Q6.2 If A1, A2, ... Ak are ____________ events, we know that Ai ∩ Aj = ∅, i ≠ j, and A1 ∪ A2 ∪ ... ∪ Ak = S.

- ◯ mutually exclusive
- ◯ exhaustive
- ◯ mutually exclusive and exhaustive
- ◯ exclusive

Q7.2 A ∩ (B ∪ C) = ___________.

- ◯ (A ∪ B) ∩ (A ∪ C)
- ◯ (A ∩ B) ∪ (A ∩ C)
- ◯ (A ∪ B) ∩ (B ∪ C)
- ◯ (A ∩ B) ∪ (A ∪ C)

Q8.2 (A ∩ B)' = __________.

- ◯ A' ∪ B'
- ◯ A' ∩ B'
- ◯ A ∩ B
- ◯ A ∪ B

Q9.2 The __________ of event A, denoted by P(A) is often called the chance of event A occurring.

Q10.2 If event B is subset of A, then ________.

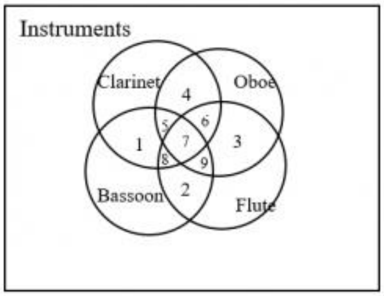
- ◯ P(A) ≥ P(B)
- ◯ P(A) ≤ P(B)
- ◯ P(A) > P(B)
- ◯ P(A) < P(B)

Q11.2 P(A ∩ B) = _____________.

- ◯ P(A) + P(B)
- ◯ P(A) + P(B) - P(A ∪ B)
- ◯ P(A ∪ B) - P(A)
- ◯ P(A ∪ B) - P(B)

Q20.2 Identify each region of Venn diagram that represents learners who play both the clarinet and oboe.

- ◯ 4
- ◯ 4+5+6
- ◯ 4+5+6+7
- ◯ 1+3+4+5+6+7+8+9



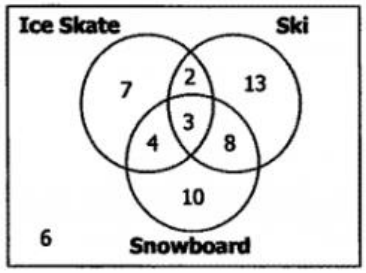
Q21.2 How many learners do not snowboard but ski?

○ 13

○ 15

○ 22

○ 28



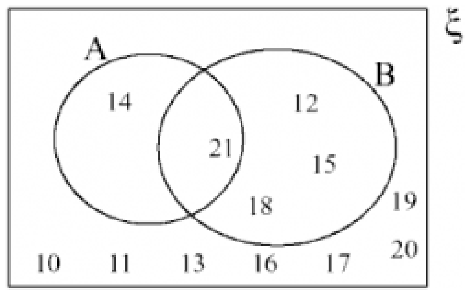
Q22.2 Which is the correct set notation for A'∩ *B*'?

○ {21}

○ {12, 14, 15, 18, 21}

○ {10, 11, 13, 16, 17, 19, 20}

○ {12, 14, 15, 18}



Q23.2 At a breakfast buffet, 23 people chose coffee and 17 people chose juice. 35 people visited the buffet. How many people chose both coffee and juice?

○ 5

○ 6

○ 17

○ 18

Q24.2 In a class of 30 learners, 19 are studying Chinese, 12 are studying Spanish and 7 are studying both Chinese and Spanish. How many learners are taking foreign languages?

○ 6

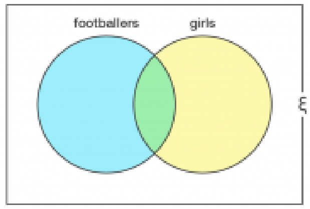
○ 12

○ 24

○ 0

Q25.2 In a class of 30 learners, 15 learners play football. 7 boys don't play football and 6 girls do play football. How many girls are there in the class?

- 9
- 16
- 14
- 23



Q30.2 A and B are mutually exclusive and exhaustive events. A has a probability 0.4. The probability of B is

- 0.06
- 0.4
- 0
- 0.6
- none of the preceding

Q31.2 A smoke-detector system consists of two parts A and B. If smoke occurs then the item A detects it with probability 0.95, the item B detects it with probability 0.98 whereas both of them detect it with probability 0.94. What is the probability that the smoke will be detected?

- 0.01
- 0.99
- 0.04
- 0.96
- none of the preceding

Q32.2 The probability that a learner passes Statistics course is 2/3 and the probability that he passes both Statistics and mathematics course is 14/45. The probability that he passes at least one course is 4/5. what is the probability that he only passes mathematics course?

- 2/15
- 4/9
- 60/135
- 112/135

Q34.2 Of a group of patients having injuries, 28% visit both a physical therapist and a chiropractor and 8% visit neither. Say that the probability of visiting a physical therapist exceeds the probability of visiting a chiropractor by 16%. What is the probability of a randomly selected person from this group visiting a chiropractor?

- 0.54
- 0.68
- 0.52
- 0.22

Q35.2 An insurance company looks at its auto insurance customers and finds that(a) all insure at least one car(b) 85% insure more than one car(c) 23% insure a sports car(d) 17% insure more than one car, including a sports car. Find the probability that a customer selected at random insures exactly one car and it is a sports car.

- 0.06
- 0.09
- 0.68
- 0.91

Q36.2 During a visit to a primary care physician's office, the probability of having neither lab work nor referral to a specialist is 0.21. Of those coming to that office, the probability of having lab work is 0.41 and the probability of having a referral is 0.53. What is the probability of having lab work but not having a referral?

- 0.79
- 0.26
- 0.15
- 0.38

Test I for Enumeration Methods

Q1.1 Suppose that an experiment (or procedure) E_1 has n_1 outcomes and, for each of these possible outcomes, an experiment (procedure) E_2 has n_2 possible outcomes. Then the composite experiment (procedure) E_1 E_2 that consists of performing first E_1 and then E_2 has ________ possible outcomes.

- $n_1 \times n_2$
- $n_1 + n_2$
- n1
- n2

Q2.1 Suppose that n positions are to be filled with n different objects. How many possible arrangements does this produce?

- n
- n × n
- n!
- n^n

Q3.1 If only r positions are to be filled with objects selected from n different objects, $r \leq n$, then the number of possible ordered arrangements is

- $\frac{n!}{(n-r)!}$
- $\frac{n!}{(n-r+1)!}$
- $\frac{n!}{r!}$
- n^r

Q4.1 If r objects are selected from a set of n objects, and if the order of selection is noted, then the selected set of r objects is called an ______ sample of size r.

- ordered
- unordered
- random
- relevant

Q5.1 ___________________________ occurs when an object is selected and then replaced before the next object is selected.

Q6.1 Compute 0! = ?

____

Q10 There are four bus lines between A and B; and three bus lines between B and C. The number of ways a person round trip by bus from A to C by way of B will be:

- 12
- 7
- 3
- 4

Q12 A woman has 5 blouses, 3 skirts, and 4 pairs of shoes. How many different outfits consisting of a blouse, a skirt, and a pair of shoes can she wear?

- ○ 12
- ○ 27
- ○ 60
- ○ 132

Q14 In designing an experiment, the researcher can often choose many different levels of the various factors in order to try to find the best combination at which to operate. As an illustration, suppose the researcher is studying a certain chemical reaction and can choose two levels of temperature, two different pressures, and two different catalysts. To consider all possible combinations, how many experiments would need to be conducted?

- ○ 6
- ○ 8
- ○ 6561
- ○ 40320

Q16 How many four-letter code words are possible using the letters in IOWA if the letters are allowed be repeated?

- ○ 4
- ○ 16
- ○ 24
- ○ 256

Q20 A special type of password consists of four different letters of the alphabet, where each letter is used only once. How many different possible passwords are there?

- ○ 426
- ○ 456,976
- ○ 14,950
- ○ 358,800

Q22 Assuming that any arrangement of letters forms a word, how many words of any length can be formed from the letters of the word SQUARE? (No repeating of letters)

- ○ 82
- ○ 720
- ○ 1,956
- ○ 9,331

Q24 The number of different permutations of the word BANANA is:

- ○ 720
- ○ 60
- ○ 120
- ○ 360

Q26 Find the number of words, with or without meaning, that can be formed with the letters of the word 'INDIA'.

- ○ 24
- ○ 60
- ○ 120
- ○ 625

Q28 In how many ways can the letters of the word APPLE can be rearranged?

____

Q30 In a colony, there are 55 members. Every member posts a greeting card to all the members. How many greeting cards were posted by them?

- ○ 990
- ○ 890
- ○ 2970
- ○ 1980

Test II for Enumeration Methods

Q1.2 Suppose that an experiment (or procedure) E_1 has n_1 outcomes and, for each of these possible outcomes, an experiment (procedure) E_2 has n_2 possible outcomes. Then the composite experiment (procedure) E_1 E_2 that consists of performing first E_1 and then E_2 has $n_1 \times n_2$ possible outcomes. What principle is described in the above passage?

- ○ Multiplication
- ○ Permutation
- ○ Combination
- ○ Branching

Q2.2 Each of the $n!$ arrangements (in a row) of n different objects is called a ___________ of the n objects.

- ○ multiplication
- ○ permutation
- ○ branch
- ○ combination

Q3.2 Each of the nPr arrangements is called a permutation of ____ objects taken ____ at a time.

- n, r
- r, n
- n, n
- r, r

Q4.2 If r objects are selected from a set of n objects, and if the order of selection is irrelevant, then the selected set of r objects is called an _______ sample of size r.

- ordered
- unordered
- relevant
- random

Q5.2 __________________________ occurs when an object is not replaced after it has been selected.

Q6.2 Compute ${}_4P_2$ = ?

Q11 A learner can take one of four Mathematics sections and one of five English sections. The number n of ways he can register for the two courses, is:

- 4
- 9
- 20
- 72

Q13 A boy found a bicycle lock for which the combination was unknown. The correct combination is a four-digit number, $d_1d_2d_3d_4$, where di, i = 1,2,3,4, is selected from 1, 2, 3, and 4. How many different lock combinations are possible with such a lock?

- ◯ 4
- ◯ 16
- ◯ 24
- ◯ 256

Q15 Suppose the license plate of a state is composed by two letters followed by a three-digit integer (leading zeros are permissible and the letters and digits can be repeated). How many different license plates are possible?

- ◯ 26*25*10*9*8
- ◯ 26*26*10*10*10
- ◯ 26*2+10*3
- ◯ (26+10)*5

Q17 A restaurant offers 5 choices of appetizer, 10 choices of main meal and 4 choices of dessert. A customer can choose to eat just one course, or two different courses, or all three courses. Assuming all choices are available, how many different possible meals does the restaurant offer?

- ◯ 329
- ◯ 129
- ◯ 200
- ◯ 19

Q21 A password consists of two letters of the alphabet followed by three digits chosen from 0 to 9. Repeats are allowed. How many different possible passwords are there?

- ○ 492,804
- ○ 650,000
- ○ 676,000
- ○ 1,757,600

Q23 In how many ways can 10 DVDs be chosen to arrange a case with slots for 3 discs?

- ○ 600
- ○ 720
- ○ 840
- ○ 1000

Q25 In how many ways can the letters in the word “Missouri” be arranged?

- ○ 5040
- ○ 10,080
- ○ 40,320
- ○ 20,160

Q27 How many different words can be formed with the letters of the word ‘SUPER’ such that the vowels always come together?

- ○ 5
- ○ 48
- ○ 60
- ○ 3125

Q29 10 learners have appeared in a test in which the top three will get a prize. How many possible ways are there to get the prize winners?

____

Q31 In Daya's bag there are 3 books of History, 4 books of Science and 2 books of Maths. In how many ways can Daya arrange the books so that all the books of same subject are together?

- ○ 9
- ○ 6
- ○ 8640
- ○ 1728

IRB Approval

Institutional Review Board
Division of Research and Innovation
Office of Research Compliance
University of Memphis
315 Admin Bldg
Memphis, TN 38152-3370

December 15, 2020

PI Name: Genghu Shi
Co-Investigators:
Advisor and/or Co-PI: Xiangen Hu
Submission Type: Initial
Title: Learning Probability by Self-explanation
IRB ID: #PRO-FY2021-201
Exempt Approval: December 14, 2020

The University of Memphis Institutional Review Board, FWA00006815, has reviewed your submission in accordance with all applicable statuses and regulations as well as ethical principles.

Approval of this project is given with the following obligations:

1. When the project is finished a completion submission is required
2. Any changes to the approved protocol requires board approval prior to implementation
3. When necessary submit an incident/adverse events for board review
4. Human subjects training is required every 2 years and is to be kept current at citiprogram.org.

For any additional questions or concerns please contact us at irb@memphis.edu or 901.678.2705

Thank you,
James P. Whelan, Ph.D.
Institutional Review Board Chair
The University of Memphis.

Institutional Review Board
Division of Research and Innovation
Office of Research Compliance
University of Memphis
315 Admin Bldg
Memphis, TN 38152-3370

June 21, 2021

PI Name: Genghu Shi
Co-Investigators:
Advisor and/or Co-PI: Xiangen Hu
Submission Type: Modification
Title: Learning Probability by Self-explanation
IRB ID: #PRO-FY2021-201
Level of Review: Full Board

Approval: June 17, 2021
Expiration: --*

The University of Memphis Institutional Review Board, FWA00006815, has reviewed your submission in accordance with all applicable statuses and regulations as well as ethical principles.

The modification is approved.

Approval of this project is given with the following obligations:

1. This IRB approval for modification has an expiration date, an approved renewal must be in effect to continue the project prior to that date. If approval is not obtained, the human subjects consent form(s) and recruiting material(s) are no longer valid and any research activities involving human subjects must stop.
2. When the project is finished a completion form must be submitted.
3. No change may be made in the approved protocol without prior board approval.
4. Human subjects training is required every 2 years and is to be kept current at citiprogram.org.

**Modifications do not extend the expiration of the original approval*

Thank you,
James P. Whelan, Ph.D.
Institutional Review Board Chair
The University of Memphis.