

Measuring effects of guided questioning support on inquiry performance in ZAPs

Tim Post

University of Twente

Master's thesis

October 23th 2009

Supervisors

Dr. T.H.S. Eysink

University of Twente

Dr. C. Hulshof

University of Utrecht

Table of Contents

1. Introduction	4
1.1 ZAP composition	4
1.2 Inquiry performance in ZAPs	5
2. Theory	6
2.1 Inquiry learning	6
2.2 Assessing inquiry performance	7
2.3 Scaffolding inquiry learning	8
2.4 Questioning as support framework	9
2.5 Guiding question stems as scaffolds	11
2.6 Hypotheses	12
3. Method	13
3.1 Participants	13
3.2 Design	13
3.3 Instruments	13
3.3.1 QFF	14
3.3.2 QSL	14
3.4 Assessment	15
3.5 Coding and scoring	16
3.6 Procedure	17
4. Results	18
4.1 Pretest	18
4.2 Posttest	18
4.3 Questions formulated	19
4.3.1 Number of questions posed	19
4.3.2 Kinds of questions posed	20
4.3.3 Strategic power of questions posed	22
4.4 Learners' experiences	23
5. Discussion	23
5.1 Unfamiliarity with guided questioning	24
5.2 Task dependency of support measure	25
5.3 Improving just-in-time support	26
5.4 Improving the question stem list	26
6. Conclusion	27
References	28

Abstract

An innovation in the field of software-based instruction are ZAPs: short and interactive computer programs, intended to encourage learners to experience and learn about psychological phenomena in an active, self-explanatory and motivational manner. A number of studies shows that inquiry learning involves many different cognitive processes that need to occur simultaneously on different levels. It is therefore perhaps not surprising that many learners experience difficulty when engaged in tasks of inquiry. In this empirical study, the effect of guided questioning support in ZAPs is measured to examine learners' improvement in inquiry performance. Two research questions are examined: (1) how does guided questioning affect learners in their use of inquiry strategies during learning performance, and (2) how the use of those strategies affect different types of knowledge constructions in ZAPs? An experiment was conducted in which inquiry performance was examined by comparing one group of learners who used guided questioning support while working with ZAPs, to two control groups who worked with the same ZAPs from which this "question-based" scaffold was either partly or fully removed. Improved inquiry learning performance was expected for learners who were allowed to work with the complete guided questioning support. Results show that guided questioning support seems to significantly impair learning performance in ZAPs and that learners are skeptical of its usefulness. It is discussed that learners are unfamiliar with guided questioning and need specialized prior instruction to familiarize themselves with utilizing questions as effective strategies for inquiry. Although results in this study show a negative relation between guided questioning support and inquiry performance, additional follow-up studies are required to fully explore and understand the extend to which guided questioning affects inquiry learning in different settings, learning tasks, and learning environments. Further research on guided questioning and the underlying inquiry learning framework used in ZAPs, could shed new light on the value of guided questioning and ZAPs as innovative learning tools in educational practices.

Keywords

Inquiry learning, guided questioning, ZAPs.

1. Introduction

In the domain of psychology, a student usually has to memorize many basic facts about many different psychological phenomena. These phenomena include among others: 'Split-brain', 'Ponzo-illusion', 'Classical conditioning', 'Gate control theory', and 'Prisoner's dilemma'. Especially in introductory courses, rote or expository learning is commonly viewed as a standard means to allow students to get acquainted with these phenomena. However, as an empirical science, students should also be tempted to go beyond the mere facts and acquire a critical way of thinking by questioning what these facts mean and how they have been derived from empirical studies. As such, studying psychology should include more authentic forms of learning in which students assume the roles of scientists by experiencing these psychological phenomena firsthand and conduct experiments to discover the underlying concepts and principles of these phenomena themselves.

Abbreviated in Dutch as "Very Interactive Psychology", ZAPs are interactive computer programs, designed for first-year psychology students to experience and learn about a wide scope of psychological phenomena in an engaging, self-explanatory and motivational manner (Hulshof, Eysink & De Jong, 2006). All ZAPs are designed from the principles of inquiry and experiential learning (Hulshof, Eysink, Loyens & De Jong, 2005; see also Kolb, 1984). This form of learning invites the learner to actively connect theory with everyday life experiences in contrast to traditional forms of rote learning where this context is absent. The ZAP project has won several awards, including the European Academic Software Award (EASA) in 2004, the National ICT Award in the category of government/non-profit in 2004 and has been bought by the American publisher Norton & Company as an accompaniment to its psychology textbooks.

1.1 ZAP composition

A ZAP consists of four basic components: (a) Introduction, (b) Activity, (c) Theory, and (d) Further Information. Although these components only seem to be linear, the learner is free to walk through the program in any order preferable (for further information, please see Hulshof et al. 2005). Figure 1 shows an example screen of a ZAP. In this particular case, the ZAP concerns the topic of classical conditioning.

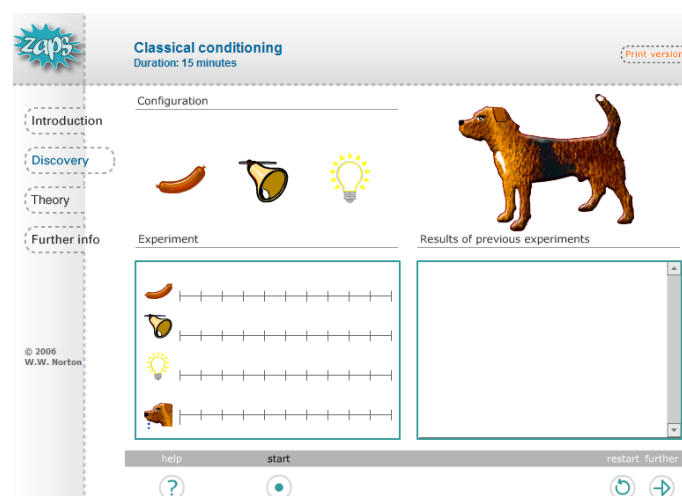


Figure 1. Example of the *Activity* component of a ZAP on classical conditioning.

The *Introduction* of a ZAP provides a concrete and realistic example out of everyday life that pertains the phenomenon. This enables learners to easily recognize their prior experiences of the phenomenon and get motivated from the initial start of the program. The *Activity* component is paramount to the overall function behind ZAP's experiential-based and inquiry-based principles. Three types of activities are concerned: (a) *experiences*, in which learners react to a set of stimuli to get personally acquainted with the phenomenon, (b) *experiments*, engage learners as participants in a simulated experiment and compares their results with standardized data from the original studies, and (c) *discoveries*, in which learners are engaged as researchers and must set up and conduct an experiment in order to investigate the underlying principles of the phenomenon themselves. Differences amongst ZAPs are mainly found in the nature of this *Activity* component, in which some ZAPs combine one or more of the three types of activities into one design. The *Theory* component explains the theoretical model of the phenomenon, previously experienced in the *Activity*, to encourage learners to reflect upon their experiences, experiments or discoveries in order to derive meaning. Lastly, the *Further Information* component entails additional descriptions that are relevant to the phenomenon derived from real-life situations.

1.2 Inquiry performance in ZAPs

By conducting usability tests involving both teachers and students, several important findings were made concerning the effectiveness and user experience of ZAPs. Foremost, learners reported to be very positive while working with ZAPs and found them motivating. The analysis of log-files recorded that most learners used all components. Teachers indicated that the user interface was easy to use and the instructions to engage in the phenomena were sufficiently clear.

An empirical study has been carried out to systematically investigate the learning effect of the *Activity* component in ZAPs (Hulshof et al. 2005). One group of first-year students worked with complete ZAPs, including all components, while a control group worked with the same ZAPs but without the *Activity* component. In the case of the control condition, the *Activity* component consisted of a mere description of an activity but no real hands-on activity was possible (i.e., experiencing, experimenting, or discovering). Surprisingly, data gathered from a pretest, posttest and retention test indicated that there were no special gains for the learners who had been able to use the *Activity* component. The study found no significant differences in learning outcomes on both the posttest and retention test that could account for the *Activity* component. Moreover, the control group showed to have even performed slightly better than the experimental group did, possibly indicating that the control group was in a stronger reading mode, as the recorded log-files showed, spending more time on the non-*Activity* components. The knowledge of the experimental group did however show a smaller decline from the time between the posttest and retention test relative to the control condition, indicating that the *Activity* component accounts for good long-term learning effects in ZAPs. Both groups showed furthermore little time spent at each component, little to no revisiting of experiments after reading the *Theory* component, and overall fair to poor performance on posttests and retention tests (Eysink, Hulshof, & Loyens, 2004). Whether the *Activity* component was there or not did not seem to matter much, in both cases learners seemed to delve very little into the learning material that was either explicitly or implicitly available. Hulshof, Eysink and De Jong (2006) argue that a better understanding of the underlying

design and assessment principles of the inquiry learning framework used in ZAPs could shed new light on the added value of ZAPs as innovative learning tools in educational practices and could encourage future research on the domain of experiential and inquiry learning.

The current study aims to revisit research on the effectivity of ZAPs by examining its underlying conceptual framework and to investigate ways to effectively scaffold learners in their application of inquiry strategies while working with ZAPs. In particular, it is reasoned that guided questioning might foster learners in their application of inquiry strategies during learning, and how this improved inquiry performance could attribute to different types of knowledge constructions.

2. Theory

To conceptualize the underlying framework used in ZAPs, a thorough description of its instructional design principles is relevant. In this section, a concise elaboration is made on the topics of inquiry learning to conceptualize underlying transformative and regulative learning processes, what kind of scaffolding is most appropriate in context of learners working with ZAPs, and conclusively how guided questioning might be appropriate and effective to scaffold inquiry performance in ZAPs.

2.1 Inquiry learning

Recent computer technologies have made it possible to provide learning environments in which inquiry activities are embedded and supported with dedicated tools. De Jong and Van Joolingen (1998) refer to this computerized form of experiential learning as “discovery learning” in which learners are required to devise one or more hypotheses and validate those by performing a number of relevant experiments. Learners are required to assume the role of scientists to “discover” the (either explicit or implicit) content of a domain themselves. According to Zimmerman (2000), the scientific discovery process in general includes both reasoning and problem-solving skills with the ultimate goal of generating and then appraising the tenability of a hypothesis. The extent to which discovery learning environments provide scaffolds to guide the learner through this process of discovery, forms the basis on which current research on inquiry learning is focused. Njoo and De Jong (1993) elaborate on inquiry by addressing two main processes that can be distinguished from an information-processing description: (a) transformative processes, that aim to produce new information (i.e., general strategies implicated in experimental design and evidence evaluation) and (b) regulative processes, that involve the control of one's own inquiry learning process on a metacognitive level (i.e., general strategies implicated in securing commitment and focus during task performance). Both processes need to occur simultaneously in order for effective inquiry learning to take place.

Klahr and Dunbar (1988) propose the SDDS-model that aims to conceptualize scientific reasoning processes in context of discovery learning. This model argues for three particular stages of inquiry: (a) a hypothesis phase, (b) experiment phase, and (c) an evaluation phase. Since Hulshof et al. (2005) argue that ZAPs are constructed upon principles derived from discovery learning, the SDDS-model seems an appropriate framework to utilize for the design of instructional support in this study. During the hypothesis phase, learners are required to devise hypotheses that aim to explain the concepts and relationships pertaining a domain. During the subsequent experiment phase these hypotheses are tested by devising experiments

or by observing, and by doing so, gathering and analyzing data in order to validate the hypotheses articulated. A hypothesis can be tested either by seeking evidence in accordance with its prediction, or by devising experiments that aim to reject its prediction (Zimmerman, 2000). Finally, during the evaluation phase, all evidence is collected and organized in order to either approve, reject or refine hypotheses stated and potentially the whole discovery process starts all over. As such, inquiry can be perceived as an iterative process of sense-making implicated in both transformative and regulative processes.

2.2 Assessing inquiry performance

A pertinent issue underlying the nature of ZAPs and the kind of conceptual framework it is designed upon, is the kind of knowledge that results and the methods that are used to measure the effects of inquiry learning. In ZAPs, inquiry learning outcome was measured by distinguishing between knowledge and insight questions (Hulshof et al. 2005). Knowledge questions would merely concern reproducing facts, while insight questions require learners to model a particular psychological phenomenon onto different situations. It was expected that the gain in factual knowledge would be the same for both conditions, while the insight in the psychological phenomena would be higher for learning with complete ZAPs compared to learning with ZAPs from which the *Activity* component was removed. Based on the surprising test results that the control condition slightly outperformed the experimental one, it was argued whether the insight questions really tapped real insight. Most of the insight questions involved imagining a situation, which the control group that was only provided the textual components of ZAPs, could likely do as easily based on their textual information. Therefore, it is paramount that any knowledge test used to assess learning outcome in ZAPs covers both 'explicit' and 'implicit' knowledge. Although this study is primarily concerned with supporting and fostering the inquiry learning processes of ZAPs, understanding the methods used to assess the effects of inquiry learning is essential for making any grounded statements on any support measure.

The notion that Swaak and De Jong (2001b) use as their introduction to the assessment of inquiry learning, is the statement of Thomas and Hooper (1991, p. 479): "the effects of simulations are not revealed by tests of knowledge." Merely addressing factual knowledge as an indication for inquiry learning performance is not sufficient and does not cover the tacit and intuitive nature of inquiry learning outcome. Importantly, studies have shown that the learning outcome of inquiry learning is insight and deep understanding, not necessarily that it results in more knowledge (e.g., Swaak & De Jong, 1996). Unfortunately, no single method currently exists that explicitly measures the core gains of inquiry learning. Therefore, currently the best approach for measuring inquiry learning outcome is a pragmatic one, one that is also carried out by Reid, Zhang and Chen (2003) who adopt four different tests that respectively assess four different kinds of knowledge: (a) *principle knowledge*, entailing seven multiple-choice items on the general principle about the factor(s) that can pertain a particular phenomenon, (b) *intuitive understanding*, in which five multiple-choice items are used in support with pictures of situations that ask students to make grounded predictions, (c) *flexible application*, eight multiple-choice items to determine how learners are able to transfer their acquired knowledge to new situations, and (d) *integration of knowledge*, where learners are asked to what extent the presented situations or concepts are related to the phenomenon that is specifically dealt with. In addition to these four features, the measurement of *factual knowledge* could be added to make the test domain more complete.

2.3 Scaffolding inquiry learning

Inquiry learning involves many different cognitive processes that need to occur simultaneously on different levels (Hulshof & De Jong, 2006). Thus, it is perhaps not surprising that many learners experience difficulties when engaged in tasks of inquiry. De Jong and Van Joolingen (1998) classify these difficulties of inquiry that learners experience into four different categories: (a) hypothesis generation, (b) design of experiments, (c) interpretation of data, and (d) regulation of learning. As such, there is increasing evidence that effective methods of promoting constructivist learning involve instructional guidance rather than pure discovery, structured focus rather than unstructured exploration (Mayer, 2004; Swaak & De Jong, 2001a). However, the issue of designing effective ways to scaffold inquiry learning still remains a challenging task for instructional designers. Learners need a sufficient amount of opportunity to become personally engaged in the process of sense-making while at the same time being provided enough guidance to cope with the amount of cognitive complexity presented.

An important notion is one that Mayer (2004) adds to this issue, in which many constructivist learning environments fail to address the appropriate cognitive learning processes and mistake it for implementing mere cognitive activity as ends in themselves. Merely requiring learners to engage on a cognitive activity does not necessarily 'activate' the appropriate cognitive *processes* that are required for inquiry learning to take place effectively. It could be this common misunderstanding that attributes to the many overloaded inquiry learning environments that are around today that bombard learners with dozens of "support" tools that really only make the learning task more dense and ineffective. Simply adding more support tools does not automatically yield higher levels of cognitive processing and inquiry performance. As Swaak and De Jong (2001a) argue, support measures in inquiry learning could potentially confront learners with an *extra* task rather than relieving them from cognitive overload. Mayer (2004, p. 17) concludes justifiably: "Methods that rely on doing or discussing should be judged not on how much doing or discussing is involved but rather on the degree to which they promote appropriate cognitive processing."

The basic notion of scaffolding inquiry as a framework for investigations into cognitive processes and instructional design, is to provide support in such a way that learners are able to cope with large and complex amounts of information: designing instructional material that "fits in" the cognitive architecture of the learner's mind (Paas, Renkl & Sweller, 2003). Van Merriënboer, Kirschner and Kester (2003) explain that successful scaffolding enables the learner to achieve a goal or action not achievable without that support. Additionally, when the learner achieves the desired goal, this support gradually diminishes or fades until it is no longer needed. Scaffolding allows learners to assume responsibility of their own problem-solving and sense-making process, and unobtrusively alleviates any cognitive overload that prevents the appropriate (cognitive) learning processes to occur.

An important issue that underlies the research of scaffolding, is the concept of 'just-in-time' information; not only providing learners the support they need, but additionally *when* they need it (Hulshof & De Jong, 2006). Next to the kind of support in transformative processes in which learners are activated to apply strategies for inquiry, learners also need to be supported in their regulative processes to access various tools and to secure commitment to their investigation (Njoo & De Jong, 1993). This support should be as unobtrusive to the learner as possible. This would enable a particular opportunity for scaffolding in which the learner can access support whenever he or she feels it is needed, and can gradually diminish access when the inquiry process becomes more familiar and habitual. Although this support

requires a certain amount of self-monitoring and responsibility on the side of the learner, it provides a natural way to call for structured help whenever the learner feels it is needed.

2.4 Questioning as support framework

In order for learning to occur, experiences must be abstracted, related and incorporated into existing knowledge structures of the learner (i.e., prior knowledge). This abstract conceptualization and incorporation of the experience is essential for it to be considered a *learning* experience. A key metacognitive strategy that enables learners to abstract these experiences and construct new knowledge, is the ability to reflect on the investigations one undertakes while exploring and learning about a particular domain. As Klahr and Dunbar (1988) propose the SDDS-model to conceptually structure these explorations, metacognition is promoted as the hallmark of effective inquiry learning. By hypothesizing, experimenting, and evaluating, new concepts and principles are systematically derived and incorporated onto the learner's existing knowledge structures to form an improved understanding of a phenomenon. Since new knowledge is constructed upon prior knowledge and prior experiences of the learner, inquiry learning by its very nature is a personal form of sense-making (e.g., Bruner, 1961; Kolb, 1984). Reflective inquiry learning could therefore be defined as a process of constructing knowledge by continuously resolving cognitive conflict between newly acquired information that is incorporated into existing mental structures of the learner, to predict the outcome of a certain principle in different situations. The nature of reflection thus strongly promotes the subjectivity of inquiry. For learning to yield any significance to the learner, inquiry learning must offer appropriate scaffolds by which knowledge can be constructed upon personal and cognitively rich experiences.

The learner's ability to successfully cope with complex tasks of inquiry, highly depends upon the learning environment's design to activate, regulate and transform reflective learning processes. Pertaining the previously stated definition of reflective inquiry learning, scaffolding reflection learning could be operationalized by aiming to support the following attributes: (a) perplexity (i.e., cognitive conflict), (b) prior knowledge (i.e., existing mental structures) and (c) predictability. A particular cognitive process that captures the essence of these three constructs of reflective inquiry learning, is students' spontaneous questioning. A study of Van der Meij (1994) concerns a componential analysis on spontaneous student questioning, where questioning is seen as a corner stone of reflection. Starting from personal questions that arise from experiencing a conflicting phenomenon, learners are supported to articulate their conflict awareness in order to find the right words for constructing question sentences and to search for information by using their questions as personal strategies for inquiry. Spontaneous student questioning constitutes three phases: (a) a phase of puzzlement or cognitive conflict in which conflicting concepts are experienced, (b) a phase of question formulation in which the learner needs to find the right words and structure to compose a question, and (c) a phase in which information is sought to answer the question stated. Although these consecutive phases seem to be linear, in practice they are often part of an ongoing, intuitive and iterative process of questioning and re-reading information transcripts many times over in order to gain deeper understanding. Only when the learner finds something strange or conflicting, a reflective attitude towards learning is assumed.

Hulshof and De Jong (2006) argue that many studies show that prior domain-specific knowledge plays a vital role in determining the application of the processes that learners undertake while experimenting and constructing new knowledge (e.g., see also De Jong & Van Joolingen, 1998). Based on this fact, in their study, learners are provided with the

opportunity to benefit from 'knowledge tips' in order to increase knowledge, foster their ability to orientate their inquiry learning process (because they should now be able to separate relevant from irrelevant variables), and what should then result in improved knowledge construction. Their results do show a better knowledge gain for the experimental group that was able to access these 'knowledge tips' in contrast to the control group that could not access these tips (test domain did not cover the specific content contained within these knowledge tips). However, mixed results were found in the gain in discovery skills, where both groups showed a relatively poor performance (only a few domain-specific principles were derived by both groups). One possible explanation for this result, based on the model of student questioning, is that providing learners 'knowledge tips' (even though these tips contain both domain concepts as well as useful strategies) is just another way of direct instruction in which otherwise valuable 'implicit' concepts and strategies are now offered 'explicitly'. If it is assumed that questioning is indeed the "root" form of reflective inquiry behavior, than this reasoning actually suggests that providing learners with more knowledge by direct instruction, diminishes the likelihood of learners to articulate knowledge gaps into strategic questions for delving deeper into the domain to *discover* these 'missing links' themselves. As such, learners are not tempted to apply any discovery skills. Simply improving one's knowledge does not necessarily mean that an inquisitive attitude towards a domain is assumed. Learners should rather be involved into the articulation of their own perplexities and being provided the help to articulate those cognitive conflicts into operational strategies to explore a domain more thoroughly. For support measures to yield proper knowledge construction, any kind of 'tip' should be thought-*provoking* rather than thought-providing. Although intimately bound to each other, the trigger for inquiry learning is questioning, not answering.

In this study, based on the same fact that prior knowledge is paramount to the underlying processes of inquiry learning that learners undertake, learners should rather be provided with 'question tips' that support the articulation process of their experience of cognitive conflict by eliciting questions that are *conflicting* with their prior knowledge. By doing so these questions become strategies for uncovering the domain based on the learner's own knowledge and curiosities. Then, as more knowledge is acquired, these questions are likely to become even more appropriate (because now relevant questions can be separated from irrelevant ones) and subsequently inquiry performance becomes more effective.

The model of student questioning helps learners in particular to find the right words to articulate their cognitive conflict into operational questions. However, merely requiring learners to articulate their perplexities does not necessarily guarantee that those perplexities will guide them through all concepts and principles pertaining a domain. One could hypothesize on a particular situation in which a learner (unwillingly) keeps 'lingering' in the same kind of conflict and thus is unable to delve any deeper or further into the learning material (i.e., as such the learner feels he or she is "stuck" or mistakingly perceives to have covered the whole domain without really having done so). Therefore the learner must not only be stimulated to ask questions but in addition also be supported in asking the *right* kind of questions to explore and understand a domain more fully.

Tabak and Reiser (2008) argue that this way of thinking is like the idea of cultivating a "disciplinary stance" in which learners are involved with a propensity to focus on particular questions, particular concerns and particular reasonings related to the domain that is dealt with. Technology can in addition play an important role in cultivating this disciplinary stance by promoting the raising of certain questions, investigation methods, data-analyses, and ways of explaining, that reflect disciplinary values and principles of scientists in that field. It

requires learners to activate prior knowledge, confront presumptions, and start using domain-dependent questions to explain phenomena in an authentic and systematic manner. It can be argued that a scientist is not someone who is primarily competent in providing answers, but is rather someone who is capable of posing the right kind of questions. Asking questions such as 'How does this relate to that? What causes X to effect Y? What would happen if I experiment with variable X? What is the evidence in support of this? How could I falsify my conclusions? Does XY also apply to AB?' and using those questions as strategies to investigate and explain a particular phenomenon. From the view of ZAPs, not merely providing "senseless" domain-generic support (e.g., monitoring tools, hypothesis scratchpads, planning tools, and process coordinators), but providing learners cognitive strategies that are in accordance with the "sense-making" methodologies of psychologists in the field. As such, the learner adopts (and potentially internalizes) the sense-making methodology that the inquiry environment represents.

2.5 Guiding question stems as scaffolds

Exhaustive research by King (e.g., 1991, 1992, 1994, 1995; see also Martino & Maher, 1999) in context of expository learning and problem solving shows that if learners are taught to articulate more improved questions, they subsequently improve their thinking capabilities and consequently have better learning outcomes. Research on *guided* questioning aims to provide pre-structured sets of questions by which learners work to investigate a particular domain. Similar to Van der Meij, King (1995) advocates that an important feature of inquiry learning is that learners are not merely searching for answers to the instructor's questions, but rather pose and answer questions that originate from their own interests, lack of understanding, and experience of conflict awareness. By providing general exemplarily question stems as structured guides, learners generate their own effective and relevant thought-provoking questions to delve deeper into the learning material (e.g., 'What would happen if ...?', 'What is another way to look at ...?', 'How does ... affect ...?', 'Now that I know about ... should I ...?', 'What is analogous to ...?', etc.). Question stems guide learners through a metacognitive process of articulating their own cognitive conflict and adopting to strategically pre-articulated question structures. By adapting to the generic question stems, learners fill in the blanks with *specific* content relative to their existing knowledge structures, perplexities and search interests (King, 1995). Because the question stems control the quality of the questions to be articulated, they so indirectly shape the answers to those questions. In her research, King (1990, 1992) demonstrates that when learners work with these question stems, learning is markedly enhanced. It suggested that learners internalize experience-based questions (rather than lesson-based ones) and are able to apply it to new tasks (see also Xun & Land, 2004). It is postulated that different types of guiding questions might promote the building of qualitatively different types of knowledge structures.

These results seem to be contrasted however by a study of Wilhelm and Beishuizen (2004) who have examined to what extent think-aloud protocols as research methods might accidentally scaffold learning and thus influence the learning processes one wants to study. The authors base their hypothesis on research done by Klahr and Carver (1995) who argue that the extra task of stimulating learners to verbalize their thought processes during task performance is likely to increase learning outcome. Questions like 'What are you going to find out? What do you think the outcome of this experiment will be? What have you found out?' present an implicit underlying investigation systematicity that could potentially scaffold learning and guide learning in the same way as guided questioning in context of this study

proposes. Their study concluded that asking standardized questions during task performance did actually not influence learning outcome. No differences were found except that learners in the no questioning condition repeated experiments more often. Wilhelm and Beishuizen (2004) reason that these findings might be task-specific and prone to the limited amount of participants that were used in their study. However, reasoning from studies mentioned earlier by Van der Meij (1990, 1994) and King (1991 – 1995), there could be other interesting views to explain the lack of learning outcome in this matter. For one, reasoning on the basis of student questioning, in the study of Wilhelm and Beishuizen (2004) learners are not allowed the opportunity to formulate their own questions but are rather provided generic questions asked by the instructor. As such, questions do not “spring” from the learner’s experience of perplexity and prior knowledge when engaged in the learning task and could thus not afford the activation of proper cognitive processing. Secondly, findings of Van der Meij (1994) show that question asking behavior could be influenced by social factors which in the case of think-aloud protocol might be hampered by the human presence of the instructor in the room in context of a laboratory experiment. Additionally, research done by King (1994) shows that question *stems*, rather than fully articulated questions, provide learners strategic ways in which to fill in the blanks themselves with concepts related to their personal interest and evolving knowledge structures during task performance. So doing, building upon the fact that question asking behavior is highly subjective and more effective when articulated by the learner rather than by the instructor.

ZAPs are intended as self-directed modules for learners to investigate on a wide variety of psychological phenomena individually. King (1995), on the other hand, employs a form of reciprocal peer questioning in which learners formulate their questions based on question stems and consequently use their questions in peer groups for articulating and answering their questions. Basing claims on Webb’s (1989) extensive research on interaction and learning in peer groups, guided questioning seems to be somewhat more profitable for learners who provide explanations to others in group work rather than in context of individual student use (King, 1992). It is reasoned that when learners work in small groups, the sum of prior knowledge and experiences might be more differentiated in body, likely resulting into more rich cognitive conflict. Resolving that conflict results in a process of questioning and answering where knowledge is constructed through discussion and thought-provoking questioning. From the viewpoint of learner autonomy, King (1992) explains that the individual use of guided questioning is beneficial over subjects using no guided support, but that reciprocal peer questioning remains a slightly more effective form of guided questioning.

Based on findings of both Van der Meij and King, it might be reasonable to investigate the usefulness of guided questioning in context of inquiry learning as an effective means to support and improve inquiry performance of learners working with ZAPs.

2.6 Hypotheses

It seems to be a valuable avenue for research to examine how guided questioning affects learners in their use of inquiry strategies during task performance and how the use of those strategies affect different types of knowledge constructions in ZAPs. Consistent with King (1991 – 1995), it is assumed that a particular support measure designed for guided questioning (i.e., question stems) as supplemental support for learners working with ZAPs, would yield improved learning performance over subjects who are unable to access this support. It is postulated that guided questioners, opposed to unguided questioners, show a larger variety of different inquiry strategies while working with ZAPs. This increased inquiry performance is

expected to show in improved posttest scores on different kinds of knowledge structures. As King (1990, 1992) postulates: when learners work with question stems, learning will not only be enhanced, but also that the use of different types of guiding questions might promote the building of qualitatively different types of knowledge structures.

3. Method

3.1 Participants

In total, 58 first- and second-year psychology students participated in the experiment: there were 24 males and 34 females with a mean age of 21 years ($SD = 2.5$). Forty-two participants were Dutch while sixteen were German. All German participants had sufficient command of the Dutch language to be able to understand the verbal instructions and written materials. Participation was scheduled as part of education (i.e., received credits) in which they volunteered for the experiment.

3.2 Design

A randomized three-group pre-post test design was used for the experiment. The pretest that was included was constructed consistent with the posttest to make comparison of results possible. Both tests were piloted. Participants were asked to study two ZAPs and assigned to one of either three conditions: a) an experimental condition consisting of “guided questioners” ($n = 18$) in which participants were required to pose questions in support with question stems, b) a second experimental condition consisting of “unguided questioners” ($n = 20$) who were required to pose questions like the “guided questioners” but were not provided any question stems to support the process of questioning, and c) a control condition, consisting of participants from which the “question-based” scaffold was entirely removed ($n = 20$). Participants were randomly assigned to all three conditions and were neither made aware of the fact that they were assigned to a particular condition nor that there were differences between the conditions. To control for sequence effects or fatigue, the sequence in which participants worked with ZAPs differed from participant to participant. It was not possible for them to work through both ZAPs in any other sequence than the one that was offered.

The main difference between the three conditions was the presence of a ‘question formulation form’ (QFF) and a ‘question stem list’ (QSL) as two supportive measures that participants were required to utilize to foster their inquiry performance (see paragraph 3.3 Instruments for further details). The control condition was not provided any support other than the ZAPs. The “unguided questioners” were provided only QFFs in addition to the ZAPs. The “guided questioners” were, like the “unguided questioners”, also provided QFFs, but in addition also handed a QSL to utilize as a supplement to articulate (more effective) questions on their QFFs. The textual length of the instructions were kept similar for both the “unguided” and “guided” questioners conditions. Due to technical circumstances, all instruction and support was provided on paper.

3.3 Instruments

Two ZAPs were selected for the present experiment. The selected ZAPs were representative of specifically the *discovery* type of ZAPs, where participants are required to assume the role

of researchers and must set up and conduct experiments in order to derive the underlying principles of a psychological phenomenon themselves. This was explicitly decided, since the proposed framework of guided questioning in this study is based upon fostering inquiry learning. The *discovery* type of ZAPs are especially classified as the most authentic inquiry learning environment amongst the *experience*, *experiment*, and *discovery* types. Hence it is assumed that *discovery* ZAPs contain more complex and tacit knowledge to be discovered than the *experiment* and *experience* types of ZAPs offer, and therefore are most suited to investigate the effectivity of guided questioning support in context of inquiry learning. The following *discovery* types of ZAPs were selected based on careful piloting to control for any initial prior knowledge that participants could have on the subjects: (1) Prisoner's dilemma and (2) Gate control theory.

3.3.1 QFF

A QFF consisted of three consecutive parts, in which a user first had room to articulate a question, a second part in which there was room to provide a provisional answer to the posed question by utilizing any prior knowledge the participant might have on the topic, and a third part in which the participant is required to provide a final (sought or investigated) answer. As such, every attempt of inquiry by the participant is facilitated by the requirement to make use of a QFF in a way that is consistent with the timing of reflection that Lin et al. (1999) propose: reflection *before*, *during*, and *after* tasks of inquiry. After the participant had finished studying both ZAPs, a handful of used QFFs were collected that “journalled” personal investigations in both ZAPs by an iterative process of questioning, provisional answering, and final answering. Learners were free to use as many QFFs that they felt they needed to explore each domain, but were explicitly told to be required to make use of the QFFs to study each ZAP.

3.3.2 QSL

The QSL constituted 30 functional and uniquely formulated question stems, characterized by both a unique inquiry strategy (i.e., inferring, validating, relating, generalizing, predicting, etc.) and one of the three inquiry phases that Klahr and Dunbar (1988) propose (i.e., hypothesizing, experimenting, and evaluating). All inquiry strategies are categorized according to the classification that Njoo and De Jong (1993) argue in terms of transformative and regulative inquiry strategies. The resulting classification is presented in Table 1 below. Opposed to the presentation of Table 1, users were presented a plain list of all 30 question stems rather than a matrix or table of some sort. This was decided to keep the design of the support measure as similar to earlier work of King (1991 – 1995) as possible to make learners' experiences comparable. In this list, all 30 question stems were randomly presented within their related inquiry phase but without mentioning their characteristic inquiry strategy. The “guided questioners” were provided a QSL as a supplement to their QFFs to help articulate more strategic and systematic questions. While “unguided questioners” were free to articulate their own questions on their QFFs, “guided questioners” were explicitly required to pick a general question stem from the QSL and use it on their QFFs. All 30 question stems depicted in Table 1 were piloted with several students prior to the experiment to make sure each stem would make sense to its users.

Table 1

'Question stem list' composed of 30 question stems, categorized into inquiry strategies and sequenced by the three stages of inquiry (SDDS-model).

Inquiry strategy		Stages of inquiry		
		<i>Hypothesize</i>	<i>Experiment</i>	<i>Evaluate</i>
<u>Transform</u>	<i>Problematizing</i>	1. What could ... be about?	2. Why does ... happen, when ... ?	3. Do I now know exactly what ... is about?
	<i>Predicting</i>	4. Could ... be about ... ?	5. What would happen if ... ?	6. Would I know the answer to the question if ... ?
	<i>Relating</i>	7. Could ... be related to ... ?	8. Does this result mean that ... is related to ... ?	9. Do I now know for sure that ... is related to ... ?
	<i>Inferring</i>	10. If ..., would ... result into ... ?	11. Is this result caused by ... ?	12. So ... influences ... ?
	<i>Generalizing</i>	13. Could it be that if ... , then ... also ... ?	14. If ... effects ..., would varying ... also lead to ... ?	15. Could I now also state that ... ?
	<i>Validating</i>	16. If ... , would I then know for certain that ... ?	17. Does this result show me that ... ?	18. Does ... mean that I now know for sure that ... ?
<u>Regulate</u>	<i>Focusing</i>	19. Would ... be the most important aspect to study?	20. Is ... the most important variable to experiment with?	21. Does this mean that ... is important to investigate further?
	<i>Searching</i>	22. Where could I learn more about ... ?	23. What is a different way to find an answer to ... ?	24. Could I read about my results in the ... section?
	<i>Planning</i>	25. Should I first investigate ..., or ... ?	26. Should I also learn how ... works, if I now know that ... ?	27. Could I learn just some more about ... ?
	<i>Organizing</i>	28. What questions should I pose first, if I like to know if ... ?	29. What series of experiments should I conduct to falsify ... ?	30. Do I now have a clear understanding of how ... works?

3.4 Assessment

Both the pretest and posttest consisted of 14 multiple-choice questions (5 alternatives) for each ZAP, adding up to a total of 28 test items per test. Since it is expected that especially first-year psychology students enrolled in the experiment would only have a limited amount of prior knowledge on both psychological phenomena covered in the selected ZAPs, it was

specifically stated in the instructions that it would not matter if they did not know the answers to the questions stated. Specifically in this case, a fifth alternative was added (i.e., “e. I don’t know”) on all test items in both tests that participants were allowed to pick if they could only guess the right answer. This was done to prevent participants from becoming initially frustrated or unmotivated prior to the experiment.

Learning effects were measured by assessing five different types of knowledge structures for each of the two ZAP domains: (a) *principle knowledge*, entailing 4 multiple-choice items on the general principle about the factor(s) that pertain both psychological phenomena, (b) *intuitive understanding*, in which 2 multiple-choice items were used in support with pictures of situations that ask students to make grounded predictions, (c) *flexible application*, 2 multiple-choice items to determine how learners were able to transfer their acquired knowledge to new situations, (d) *integration of knowledge*, 3 multiple-choice items where learners were asked to what extend the presented situation or concepts are related to the phenomenon that is specifically dealt with, and (e) the measurement of 3 multiple-choice items on (e) *factual knowledge* (e.g., concepts, definitions, etc.) was added to make the test domain more complete.

After the posttest was administered, a small questionnaire consisting of five 5-scale Likert items was provided in which each participant was required to rate his or her experience related to their time working with ZAPs. Each item concerned a virtual statement of a learner relating to a particular aspect of the experiment in which the participant needed to score to what extend he or she felt similar (i.e., 1. Not at all, 2. A little, 3. Reasonably, 4. Fairly, 5. Very much so). The following five statements were presented (illustrating differences for each condition that made each statement applicable to one of the three conditions):

1. During my time studying both ZAPs, I felt supported [control group: n.a.; “unguided questioners”: by the ‘question formulation forms’.; “guided questioners”: by the ‘question stem list’.];
2. [control group: n.a.; “unguided questioners”: By using the ‘question formulation forms’.; “guided questioners”: By using the ‘question stem list’.] I felt I was not allowed enough room to explore both ZAPs to my own accord;
3. [control group: n.a.; “unguided questioners”: By using the ‘question formulation forms’.; “guided questioners”: By using the ‘question stem list’.] I felt I was not allowed enough time to explore both ZAPs fully;
4. I felt that [control group: I did not need any extra support, “unguided questioners”: working with the ‘question formulation forms’ was unnecessary, “guided questioners”: working with the ‘question stem list’ was unnecessary];
5. I felt that [control group: I guided my investigation of each ZAP by using my own questions in mind, “unguided questioners”: the ‘question formulation forms’ were obtrusive to my learning, “guided questioners”: the ‘question stem list’ was obtrusive to my learning].

3.5 Coding and scoring

Two kinds of data-sets were coded and scored. The first kind were the answers on the pretest and posttest. Both the pretest and posttest consisted of 14 multiple-choice items for each ZAP, that could be scored with 1 point for each correct answer. This added up to a total of 28 points per test that a participant could potentially earn.

The second kind of data was derived from the QFFs that participants in both the “unguided” and “guided” conditions had used to investigate the domains of both ZAPs. Each QFF contained three constructs: a) the question articulated, b) provisional answer provided,

and c) the final answer given. For “unguided questioners”, the questions articulated were classified by using the ‘question stem matrix’ (see Table 1). Questions articulated by the “guided questioners” did not need to be classified by raters, since participants were explicitly required to pick one of the question stems from the QSL, note down its label number on their QFFs, and so doing provided the classification of each articulated question themselves. Provisional answers and final answers that learners had given, were used as qualitative measures to score each posed question on “strategic power”. An incorrect provisional answer and correct final answer would score a maximum of 2 points each and add up to a maximum total score of 4 points that a posed question could earn (1 point was assigned when an answer was only partly correct). As such, next to the kind of question that the learner had posed, the strategic power of that question was scored. This strategic power was defined as: a question is most strategic (i.e., of high quality) when a participant does not have any prior knowledge of the possible answer (this validates that the participant truly tries to bridge an actual ‘knowledge gap’, and thus controls for questions that participants could pose that merely confirm what they already know), but does succeed in finding or deriving the correct answer eventually (this validates that the posed question did yield knowledge construction successfully). In other words, questions scored with 4 points justify that the participant truly learned something new by posing this particular question, and as such, were “strategically effective”. Two raters scored the QFFs of 25% of the “unguided questioners” group. Interrated agreement was 0.91 (Cohen’s k).

3.6 Procedure

Participants were provided a personal login code and were required to fill in personal information like their name, age and college year. Then, participants were given the pretest and because it was assumed that some participants had less to no prior knowledge on the particular topics covered in the ZAPs offered, it was made explicit that (relatively) low scores were expected and thus were of no practical concern. This had to be made clear to prevent participants to become emotionally influenced prior to working with the ZAPs and affect knowledge construction in any way.

After participants had filled in their personal information forms and pretests, each participant was given access to two *discovery* type of ZAPs (i.e., ‘Prisoner’s dilemma’ and ‘Gate control theory’) in a random order based on the login code that they were provided. To ensure that all ZAPs were walked through seriously, a ‘5-minute limit’ on study time for each ZAP was mandatory. If a student liked to proceed to the next ZAP before these five minutes were spent, a pop-up appeared telling the participant to revisit the learning content and re-engage on the task seriously.

Just before participants started working with their ZAPs, specific textual instruction was provided on paper. Participants in the control condition were not provided any support and were required to start explore each ZAP freely to their own accord. “Unguided questioners” were provided blank QFFs and required to study both ZAPs by using those forms. “Guided questioners” were like “unguided questioners” also provided QFFs, but in addition to these also provided a QSL and required to pick questions stems when articulating questions on their QFFs. For both “unguided” and “guided” questioners, it was instructed that participants could use as many QFFs as they felt they needed to fully explore each ZAP. Also it was made clear that a QFF should be used either *before*, *during*, or *after* studying the ZAP, not just prior or after having studied the ZAP. Each time the participant experienced some form of cognitive perplexity concerning learning content (e.g., while reading, experimenting, discovering, etc.),

this perplexity had to be articulated into a question and a provisional answer to that question needed to be given. In the case that a participant was unable to answer the posed question, it was allowed to skip the question and revisit it later on to try to answer it again.

Once a participant had worked through both ZAPs, he or she turned in all used QFFs of each of the ZAPs (excluding participants from the control group), and was required to start on the posttest. After the posttest was administered, a small questionnaire was provided to participants from all conditions in which they were asked to rate their experiences related to their time working with ZAPs, after which the experiment was ended. For each participant the complete experiment was scheduled as a 90 minute session.

4. Results

4.1 Pretest

As expected, the results showed that prior knowledge concerning the two ZAP domains (i.e., 'Prisoner's dilemma' and 'Gate control theory') was very poor (overall $M = 3.29$, $SD = 2.42$). All participants rated the fifth alternative "e. I don't know" on almost every test item. Because of this, it was not possible to calculate a reliable value of Cronbach's alpha. In general, pretest scores showed that participants in all three conditions did not significantly differ from one another and hence were comparable ($F_{2,55} = .26$, n.s.).

4.2 Posttest

Cronbach's alpha for test reliability of the complete posttest was .72. To measure differences in learning performance between the three conditions, the scores on the posttest are compared. A multivariate analysis of variance (MANOVA) using post hoc Bonferroni testing shows that the differences on mean scores between the three conditions were significant and large ($F_{2,55} = 12.44$, $p < .01$, partial $\eta^2 = .31$). Surprisingly, participants from the control condition, who were offered no additional support during their work with both ZAPs ($M = 17.65$, $SD = 3.23$), outperformed the "unguided questioners" ($M = 16.00$, $SD = 3.64$) and "guided questioners" ($M = 12.11$, $SD = 3.71$) in their overall posttest scores. The analysis further indicates that the mean posttest scores of the control condition in comparison to the posttest scores of the "unguided questioning" condition did not significantly differ (M difference = 1.65, $p = .43$), in contrast to the comparison of the control condition to the "guided questioning" condition (M difference = 5.54, $p < .01$), and the "unguided questioning" to the "guided questioning" condition (M difference = 3.89, $p < .01$) that did significantly differ. The same pattern of the outperformance of the control condition in respect to the "unguided" and "guided" questioners is seen in specific scores of both ZAP domains that covered the test domain. The ZAP about the topic of the prisoner's dilemma was experienced as significantly more difficult than the ZAP about gate control theory ($F_{2,55} = 12.60$, $p < .01$, partial $\eta^2 = .31$).

An additional MANOVA (post hoc Bonferroni) with subsequent ANOVAs was used to compare learning performance between the three conditions, specifically related to the five types of knowledge structures that were designed in the test domain. Results show that posttest scores differ significantly and largely between all three groups (Wilks's Lambda: $F_{2,55} = 3.82$, $p < .01$, partial $\eta^2 = .27$), and specifically on integrative knowledge ($F_{2,55} = 3.72$, $p < .$

05, partial $\eta^2 = .12$), flexible knowledge ($F_{2,55} = 5.35$, $p < .01$, partial $\eta^2 = .16$), declarative knowledge ($F_{2,55} = 8.27$, $p < .01$, partial $\eta^2 = .23$), principle knowledge ($F_{2,55} = 5.37$, $p < .01$, partial $\eta^2 = .16$), intuitive knowledge ($F_{2,55} = 5.17$, $p < .01$, partial $\eta^2 = .16$).

On average over all three conditions, participants acquired more declarative knowledge than of any other kind of knowledge structure covered in the test domain. Significant differences in knowledge structures were only found for participants in the control condition, who performed better on declarative knowledge construction than “unguided questioners” (M difference = 1.05, $p < .05$) and “guided questioners” (M difference = 1.63, $p < .01$) did. On the construction of intuitive knowledge, the “unguided questioners” showed the most improved learning performance, and significantly outperformed the control group (M difference = .82, $p < .01$).

Table 2

Posttest scores of the three conditions on the five types of knowledge structures in the test domain

Knowledge type	Condition					
	Guided questioners		Unguided questioners		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Integrative knowledge (Total of 6 points)	3.00	1.50	3.75	1.21	4.10	1.07
Flexible knowledge (Total of 4 points)	1.72	0.96	2.60	1.00	2.60	1.00
Declarative knowledge (Total of 6 points)	3.22	1.26	3.80	1.54	4.85	0.88
Principle knowledge (Total of 8 points)	2.83	1.04	3.75	1.41	2.83	1.04
Intuitive knowledge (Total of 4 points)	1.28	0.96	2.10	0.72	1.8	0.7

4.3 Questions formulated

Next to the analysis of differences in overall and specific posttest scores between groups, the number and kind of questions posed by both the “unguided questioning” and “guided questioning” conditions are examined.

4.3.1 Number of questions posed

ANOVA analysis of posed questions (i.e., the QFFs) during the experiment, shows no significant difference ($F_{1,36} = 12.53$, n.s.) in the number of questions posed between both the “unguided” ($M = 8.65$, $SD = 3.35$) and “guided” questioners ($M = 7.50$, $SD = 4.00$). The number of questions posed did not show statistical significant relations to overall posttest scores nor to scores on the five knowledge structures.

4.3.2 Kinds of questions posed

As expected, the “unguided questioners” posed only a very selective variety of functional questions (see Table 3). Almost all of the questions posed were classified as “Problematizing – Hypothesizing” questions. These characterize a propensity of “unguided questioners” to naturally pose transformative hypothesizing questions when working with ZAPs, rather than experimental, evaluating, or regulative ones. In the “unguided questioners” condition, no significant correlations were found to overall posttest scores.

In contrast to the “unguided questioners”, the “guided questioners” showed to have posed a far larger variety of functional questions. This was expected, since participants were explicitly required to pick question stems from the 30 classes of questions provided on the QSL. Table 4 shows that “guided questioners” had, in comparison to the “unguided questioners”, posed more questions that were specifically related to the inquiry phase of experimentation rather than of hypothesizing. Also in the “guided questioners” condition, no significant correlations were found to overall posttest scores. Table 3 and 4 depict the correlations of all posed questions to overall posttest scores for respectively the “unguided questioning” (Table 3) and “guided questioning” (Table 4) condition.

Table 3

Number of questions posed by “unguided questioners”, supplemented with Pearson's correlations related to overall scores on the posttest. Only posed question classes are highlighted by grey cells.

Inquiry strategy		Stages of inquiry		
		Hypothesize	Experiment	Evaluate
<u>Transform</u>	<i>Problematizing</i>	Count = 144 $r = -.22$		
	<i>Predicting</i>	Count = 2 $r = -.24$		
	<i>Relating</i>	Count = 21 $r = -.24$		
	<i>Inferring</i>	Count = 2 $r = -.26$		
	<i>Generalizing</i>			
	<i>Validating</i>			
<u>Regulate</u>	<i>Focusing</i>	Count = 1 $r = -.07$		
	<i>Searching</i>			
	<i>Planning</i>			
	<i>Organizing</i>			Count = 1 $r = .26$

Table 4

Number of questions posed by “guided questioners”, supplemented with Pearson's correlations related to overall scores on the posttest. Only posed question classes are highlighted by grey cells.

Inquiry strategy		Stages of inquiry		
		<i>Hypothesize</i>	<i>Experiment</i>	<i>Evaluate</i>
<u>Transform</u>	<i>Problematizing</i>	Count = 19 $r = -.07$	Count = 3 $r = -.06$	Count = 5 $r = .26$
	<i>Predicting</i>	Count = 9 $r = .15$	Count = 29 $r = -.17$	Count = 2 $r = .38$
	<i>Relating</i>	Count = 10 $r = .13$	Count = 2 $r = .14$	
	<i>Inferring</i>	Count = 7 $r = .16$	Count = 2 $r = -.06$	Count = 3 $r = -.30$
	<i>Generalizing</i>	Count = 1 $r = .06$	Count = 2 $r = -.06$	Count = 4 $r = -.16$
	<i>Validating</i>		Count = 6 $r = .42$	
<u>Regulate</u>	<i>Focusing</i>	Count = 4 $r = .22$	Count = 1 $r = .06$	
	<i>Searching</i>	Count = 9 $r = .08$	Count = 2 $r = .33$	Count = 3 $r = -.26$
	<i>Planning</i>	Count = 1 $r = .06$	Count = 2 $r = .14$	
	<i>Organizing</i>		Count = 1 $r = .33$	Count = 14 $r = .24$

Examining the correlations of the classes of questions posed over both conditions to overall posttest scores, shows significant correlations of question class 1 “Problematizing – Hypothesize” ($r = .331, p < .05$), class 5 “Predict – Hypothesize” ($r = -.34, p < .05$), and class 12 “Inferring – Evaluate” ($r = -.32, p < .05$). Most of which had a negative relation with overall posttest scores.

When examining correlations specifically to the five kinds of knowledge assessments designed in the posttest, no statistically significant correlations were found for any questions posed by the “unguided questioners”. However, for the “guided questioners”, statistical significance was found for positive relations of question class 8 “Relating – Experiment” to the construction of declarative knowledge ($r = .51, p < .05$), both question class 19 “Focusing – Hypothesize” ($r = .55, p < .05$) and question class 23 “Searching – Experiment” ($r = .65, p < .01$) to the construction of intuitive knowledge, and question class 29 “Organizing – Experiment” related significantly positive to the construction of declarative knowledge ($r = .50, p < .05$). Question class 24 “Searching – Evaluate” related significantly negative to the construction of integrative knowledge ($r = -.51, p < .05$).

Next to analyzing correlations of single classes of questions to posttest scores, it is also relevant to correlate groups of inquiry strategies or inquiry phases to posttest scores. Results show that no significant correlations were found for the “unguided questioners” in respect to posttest scores. Only for “guided questioners”, ‘validating questions’ are moderately significant related to positive posttest scores ($r = .42, p < .1$). Over both conditions, only ‘problematizing questions’ show to relate significantly positive to overall posttest scores ($r = .33, p < 0.05$), while ‘predicting questions’ show to relate moderately significant to negative posttest scores ($r = -.30, p < .10$). Additionally grouping all questions into either groups of transformative or regulative inquiry strategies, results into no large and statistical relations to posttest scores either. Additionally, correlating groups of questions according to their assigned inquiry phase to overall posttest scores, shows that only ‘hypothesizing questions’ correlate significantly positive to scores on the posttest ($r = .34, p < .05$).

Specifying correlations to the five knowledge structures designed in the posttest, leads to a significant positive relation of ‘organizing questions’ to the construction of integrative knowledge for “unguided questioners” ($r = .44, p = .05$). For “guided questioners”, significant positive relations were found for ‘validating questions’ to the constructions of both declarative ($r = .60, p = .01$) and intuitive knowledge ($r = .49, p = .05$). Also significant positive relations were found for ‘focusing questions’ to the construction of intuitive knowledge ($r = .49, p < .05$), and for ‘organizing questions’ to the construction of declarative knowledge ($r = .48, p < .05$). Over both conditions, ‘predictive questions’ showed to relate significantly negative to all three constructions of flexible knowledge ($r = -.31, p = .05$), principle knowledge ($r = -.34, p < .05$), and intuitive knowledge ($r = -.33, p < .05$). For ‘generalizing questions’, a significant negative relation was found on the construction of intuitive knowledge ($r = -.32, p = .05$).

4.3.3 Strategic power of questions posed

Lastly, scores on provisional answers and final answers of each posed question is examined to determine how “strategically powerful” participants were in posing questions that engaged on bridging knowledge gaps. MANOVA showed that between conditions, scores on provisional answers did not significantly differ ($F_{1,36} = .09, n.s.$). However, on differences in scores on final answers, “guided questioners” show a slight but significantly improved performance ($F_{1,36} = 4.09, p = .05$, partial $\eta^2 = .10$). On “strategic power”, both groups differ slightly and not significantly ($M = 2.53, SD = .49$). When correlating all three constructs over both conditions to overall posttest scores, scores on final answers showed a significant negative relation ($r = -.33, p < .05$). Specifying this analysis to either condition, leads to the same finding only for “unguided questioners”: a significant negative relation was found on final answer scores to overall posttest ($r = -.48, p < .05$). No statistically significant correlations were found for the “guided questioners” group.

Correlating strategic power scores to the five knowledge assessments, leads to a significant negative relation between final answer scores and the construction of both declarative ($r = -.54, p < .05$) and principle knowledge ($r = -.48, p < .05$) for “unguided questioners”. For “guided questioners”, a significant positive correlation was found for “strategic power” on the construction of principle knowledge ($r = .50, p < .05$). Over both conditions, a significant negative correlation was found for “strategic power” on the construction of declarative knowledge ($r = -.39, p < .05$).

4.4 Learners' experiences

Table 5 depicts a summary of ratings from the small questionnaire that every participant needed to fill in after turning in their posttests. These ratings concerned their experiences related to their time working with ZAPs, especially to how they might felt supported by the material provided.

Table 5

Summary of average self-rated scores on learner's experience of working with ZAPs

Question	Condition					
	Guided questioners		Unguided questioners		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1. "I felt supported"	2.11	1.02	2.75	1.21	3.60	0.60
2. "I was not allowed enough room"	3.61	1.24	3.25	1.33	1.95	1.28
3. "I was not allowed enough time"	3.11	1.32	2.35	1.31	1.25	0.55
4. "I felt that support was unnecessary"	3.72	1.23	2.95	1.42	3.60	1.00
<i>(Control: "I did not need extra support")</i>						
5. "I felt support was obtrusive to learning"	3.78	1.22	2.75	1.33	3.15	1.18
<i>(Control: "Guided my learning with questions in mind")</i>						

* Score of 1 point is "Not at all" while a score of 5 points is "Very much so"

ANOVAs show that participants showed significant differences in self-ratings on their experience working with ZAPs for statement 1 ($F_{2,55} = 11.18, p < .01$), statement 2 ($F_{2,55} = 8.95, p < .01$), and statement 3 ($F_{2,55} = 13.50, p < .01$). Differences between conditions for statement 4 and 5 were computed between the "unguided" and "guided" questioners alone, since the rated statements provided for participants in the control condition were unrelated. Only statement 5 showed significant differences ($F_{1,36} = 6.21, p < .05$) where the "guided questioners" found their support more obtrusive than the "unguided questioners".

5. Discussion

This study investigated the effects of guided questioning support on inquiry learning performance of learners working with ZAPs. It was expected that participants who utilized guided questioning support would construct more differentiated and improved knowledge structures than participants who had no or only partly access to this support.

In general, all learners showed great improvement in knowledge concerning both domains covered in the ZAPs. Initially learners had virtually no prior knowledge on both topics but acquired a good to fair amount of differentiated knowledge at the end of the experiment. In spite of this, both the analysis of posttest scores and self-ratings of learners showed undoubtedly that neither the QFFs nor the QSL combined were supportive to inquiry learning in ZAPs. Even more so, this support was perceived as being unhelpful, restrictive, and even obtrusive. In all instances, participants that were not provided any guided questioning support outperformed participants from the groups that were provided either partly or complete guided questioning support. It seems that for participants working with ZAPs, the more guided question support that was provided (i.e., QFFs and QSL), the greater the decrease of inquiry performance that was yielded. Though ineffective and contradictory to our expectations, several findings that resulted from the data-analysis presented in this study do shed new light on the attributions of guided questioning support on inquiry performance in context of ZAPs. These findings lead to suggest four strands of reasoning that attempt to explain the contrasting results in comparison to research on guided questioning by King (1992 – 1995).

5.1 Unfamiliarity with guided questioning

The first explanation is one of cognitive overload (see De Jong, in press) and attempts to justify the results of this study on the lack of prior instruction for participants to familiarize themselves with the application of (guided) questioning prior to the experiment. In respect to research of King (1992 – 1995), it could be that the contrasting findings presented in this study are not much due to the theoretical validity of guided questioning as a framework for improving inquiry learning, but more related to an implicit assumption that participants were already knowledgeable of how to utilize (guided) questions as helpful strategies for inquiry, prior to the experiment. Building upon results from especially the acquired self-ratings of participants, it might be likely that participants needed more time to familiarize themselves with the concept and practice of (guiding) questioning before studying both ZAPs in context of the experiment. As such, “unguided” and “guided” questioners did not only study both ZAPs, but also needed to learn about the support measure itself while learning about both psychological phenomena. Instead of being facilitative to inquiry, guided questioning support became an extra task. Hence, inquiry performance decreased because of cognitive overload. These experiences are elaborated by Schnotz and Rasch (2005) who implement a “practice phase” prior to these kinds of studies, to avoid that learners become cognitively overloaded by an unfamiliar support tool once the experiment starts. This line of reasoning could suggest that though questioning might still be a natural cognitive process to engage learners on tasks of inquiry, simply telling learners to utilize (guiding) questioning is not, and thus should well be accounted for by proper instructional material in future studies.

An interesting and important observation that confirms this issue, is the inconsistency of provisional and final answers provided on some QFFs of both “unguided” and “guided” questioners, in comparison to the related questions stated. In some instances, the literal semantic structure of a question articulated by a participant would clearly suggest the classification of a certain type of question (with its corresponding inquiry strategy and inquiry stage), but when examining the corresponding provisional and final answer that the participant had provided, seemed to implicitly tell that the participant was actually engaged on a different cognitive conflict than the one he or she had articulated. In other words, in a number of cases, inferred from the provisional and final answers that questioners had

provided on their QFFs, some participants seemed to be unable to articulate their cognitive conflicts into reflecting questions for inquiry. In terms of necessary metacognitive skills, the articulation of conflict awareness as questions for strategies for inquiry is not by any means a natural and familiar task, and thus likely needs additional (prior) instructional support.

5.2 Task dependency of support measure

The second explanation that might justify why guided questioning support impaired inquiry performance in this particular study, is related to a probable task dependency of the guided questioning support specifically proposed here, and thus attempts to explain the contradictory results from a quite different perspective than the first explanation of cognitive overload.

This study aimed at transferring the effectivity of guided questioning from the context of expository learning and problem solving (see King, 1992 – 1995) as a support measure to the context of inquiry learning in ZAPs. In this way, the instructional materials and support tools provided in this study were specifically designed upon the framework of inquiry learning and therefore differed from the theoretical framework used in earlier studies by King. The question stem list (QSL) used in this study consisted of an underlying framework explicitly derived from design principles acquired from literature on inquiry learning. Each articulated question stem was devised according to a typical inquiry strategy (Njoo & De Jong, 1993) and inquiry stage (Klahr & Dunbar, 1988). This was decided on the assumption that *discovery* ZAPs are promoted as inquiry learning environments. It might be this assumption that is false, and could explain why participants experienced much difficulty in applying the inquiry-based (guided) questioning support when working with ZAPs. Results from the original study and results from this current study show findings in support of this way of reasoning.

The first argument is related to research results from the original study on ZAPs by Hulshof et al. (2005) of a two-group pre-post test design in which the control group that was not provided the *Activity* component actually outperformed the experimental group that was provided a complete ZAP with the *Activity* component. Overall, posttest scores showed relative poor performance for both groups. The knowledge of the experimental group did however show a smaller decline from the time between the posttest and retention test relative to the control condition, indicating that the *Activity* component accounts for good long-term learning effects in ZAPs. Still, based on these results, it can be questioned how ZAPs could be inquiry-based if the *Activity* component (which is the only space where participants are able to set up and conduct experiments to engage on their discovery skills) does not manage to improve inquiry performance. Additionally, the second argument concerns results presented in this study that show consistent results in the relatively strong construction of declarative knowledge by participants, opposed to poor performance on integrative, principle, flexible, and intuitive knowledge constructions. In spite of the presence of an *Activity* component, it seems like ZAPs engage learners on a *reading* mode rather than a *discovery* mode. These findings could suggest that although ZAPs claim to be well-designed on the basis of inquiry learning principles, they are actually not.

Hence, if the guided questioning support devised in this study truly represents authentic inquiry-based questioning, and if that support was not prone to any extraneous cognitive overload due to a seemingly unfamiliarity of participants working with the QFFs and QSL, then the (guided) questioning support used in this study might have served as a measure to “validate” the inquiry-based nature of ZAPs. Results show that many posed questions relate negatively to posttest scores while only a very small number of question classes relate

positively. In contrast to the argument of cognitive overload, the negative correlation of utilized question stems by participants might not be due to cognitive complexity but cognitive inconsistency: the QFF and QSL are aimed to force inquiry-based thinking onto a learning environment that is (unintentionally) not capable to invite inquiry learning. If this is the case, then inquiry-based guided questioning inevitably impairs learning in context of ZAPs. In any future study it might therefore be relevant to expose inquiry-based questioning support to inquiry learning environments other than ZAPs, that have proven to engage learners on authentic inquiry tasks (e.g., “SimQuest”, Van Joolingen & De Jong, 2003; “Co-Lab”, Van Joolingen, De Jong, Lazonder, Savelsbergh & Manlove, 2005).

5.3 Improving just-in-time support

A third explanation concerns the notion that underlies the research of scaffolding, the concept of ‘just-in-time’ information: not only providing learners the support they need, but additionally when they need it (Hulshof & De Jong, 2006). In this respect, in context of this study, every attempt of inquiry by the participant was facilitated by the requirement to make use of a QFF in a way that is consistent with the timing of reflection that Lin et al. (1999) propose: *before*, *during*, and *after* inquiry tasks. Participants were required to pose a new question on a QFF every time they answered their question and experienced some sort of new cognitive conflict awareness while studying the psychological phenomena pertaining in both ZAPs. Based on research of Van der Meij (1994) on the conceptualization of spontaneous student questioning, the sequential structuring proposed by Lin et al. (1999) might not have turned out to be appropriate in respect to the nature of (guided) questioning.

As Van der Meij (1994) argues, spontaneous student questioning constitutes three phases: (a) a phase of puzzlement or cognitive conflict in which conflicting concepts are experienced, (b) a phase of question formulation in which the learner needs to find the right words and structure to compose a question, and (c) a phase in which information is sought to answer the question stated. In practice, this process is often part of an ongoing, intuitive and iterative process of questioning and re-reading information transcripts many times over in order to gain deeper understanding. This contrasts the systematic timing of reflection proposed by Lin et al. (1999) and challenges the instructional material provided to participants who were required to utilize the QFFs in this pre-structured manner. During careful piloting of the QFF prior to the experiment, users showed to have trouble regulating their questioning processes without any explicit support on timing their question articulations: some only posed questions at the beginning of each ZAP, others posed questions only after they had already explored the whole ZAP, while some even entirely forgot to pose questions. Although the sequential structuring by Lin et al. (1999) did help to make the timing of questioning more explicit and transparent to the participants, it may have compromised the inherent intuitive and highly subjective nature of the “timing” of questioning and so doing, impaired inquiry performance as a result. Therefore it might be important in any future studies to allow participants more alternative opportunities to pose questions to their own accord rather than to pre-structure the timing of questioning by instructional restrictions.

5.4 Improving the question stem list

The last argument concerns the sophistication of the QSL that was intended to activate participants to pose more strategically sound questions during their study of both ZAPs. Results showed that for the “unguided questioners”, almost all questions posed (about 84%)

were categorized as 'Problematizing – Hypothesizing' questions. Although this might suggest and confirm earlier studies by, for example De Jong and Van Joolingen (1998), that "unguided" learners tend to show to obtain a very selective application of inquiry strategies, it could also serve as critique on the sophistication of the QSL for being unable to make more specific classifications. Particularly, not all questions classified as 'Problematizing – Hypothesizing' are alike. For example, learners posed original questions like: "Why might it be interesting to make a ZAP about the Prisoner's dilemma?" or "How is it like to do this experiment?". The originality of such questions challenge the flexibility of the question stems on the QSL as a tool to classify "unguided" questions. These "original" questions might well be rooted in more complex and meaningful cognitive thinking strategies than the current QSL seems to be able to capture.

6. Conclusion

Traditional views on learning, especially within the domain of psychology, involve expository teaching methods in which knowledge acquisition is perceived as gaining factual understanding. In this study, the effectivity of ZAPs as short and interactive computer programs designed for psychology students to experience and learn about a wide scope of psychological phenomena, was investigated by using the theory of guided questioning to measure effects on inquiry performance.

In spite of the negative influence that the (guided) questioning support had on inquiry performance in this study, several cases of QFFs did show the activation of proper cognitive processing. A particular participant gave the following final answer to one of its posed questions: "Did not find an answer, but I do know what my next question will be!". Or of a similar kind: "Did not find an answer, but my question did inspire me to do some other experiments". Also, learners showed to be able to honestly self-assess learning, like for instance a "guided questioner" who posed the following question: "Do I now know what the Prisoner's dilemma is all about?", provided the provisional answer "Yes", but concluded: "No, I did not. I did not know about the different variations of dilemma's yet." Or: "Would pain avoidance be the most important thing to study in this ZAP?", provisional answer: "The introduction seems to tell so", final answer: "No it was not, knowledge of different fibers is more important!". These cases, though not plentiful, show that guided questioning did, for some, facilitate an underlying personal and systematic process of investigating ZAPs according to the theory of guided questioning.

Although results in this study show a negative relation between guided questioning support and inquiry performance in ZAPs, additional follow-up studies are required to fully explore and understand the extend to which guided questioning might affect inquiry learning in different settings, learning tasks, and learning environments. Further research on guided questioning and the underlying inquiry learning framework used in ZAPs, could shed new light on the value of guided questioning and ZAPs as innovative learning tools in educational practices.

References

- Bruner, J. S. (1961). The act of inquiry. *Harvard Educational Review*, 31, 21–32.
- De Jong, T. (2006). Technological advances in inquiry learning. *Computer simulations*, 312, 532–533.
- De Jong, T. (in press). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*.
- De Jong, T., & Van Joolingen, W. R. (1998). Scientific inquiry learning with computer simulations of conceptual domains. *Review of Educational Research*, 68, 179–201.
- Eysink, T., Hulshof, C., & Loyens, S. (2004). *Finaal evaluatie rapport: ZAP, Zeer Actieve Psychologie (D7-2)*. Retrieved from http://zap.gw.utwente.nl/_Documenten/_documenten.html on June 26th of 2008.
- Hulshof, C., & De Jong, T. (2006). Using just-in-time information to support scientific inquiry learning in a computer-based simulation. *Interactive Learning Environments*, 14(1), 79–94.
- Hulshof, C., Eysink, T., & De Jong, T. (2006). The ZAP project: designing interactive computer tools for learning psychology. *Innovations in Education and Teacher International*, 43(4), 337–351.
- Hulshof, C., Eysink, T., Loyens, S., & De Jong, T. (2005). ZAPs: Using interactive programs for learning psychology. *Interactive Learning Environments*, 13(1–2), 39–53.
- King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, 27, 664–687.
- . (1991). Effects of training in strategic questioning on children's problem-solving performance. *Journal of Educational Psychology*, 83(3), 307–317.
- . (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, 27(1), 111–126.
- . (1994). Guiding knowledge construction in the classroom: Effects on teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338–368.
- . (1995). Inquiring minds really do want to know: Using questioning to teach critical thinking. *Teaching of Psychology*, 22(1), 13–17.
- Klahr, D., & Carver, S. M. (1995). Scientific thinking about scientific thinking. In: D. Kuhn, M. Garcia-Mila, A. Zohar, & C. Anderson (Eds.). *Strategies of knowledge acquisition*. Monographs of the Society for Research in Child Development, 60 (4, Serial No. 245).
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1–48.
- Kolb, D. A. (1984). *Experiential learning*. Englewood Cliffs, NJ: Prentice-Hall.
- Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J. (1999). Designing technology to support reflection. *Educational Technology, Research and Development*, 47(3), 43–62.
- Martino, A. M., & Maher, C. A. (1999). Teacher questioning to promote justification and generalization in mathematics: What research practice has taught us. *Journal of Mathematical Behavior*, 18(1), 53–78.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure inquiry learning? The case for guided methods of instruction. *American Psychologist*, 59(1), 14–19.
- Njoo, M. K. H., & De Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, 30, 821–844.

- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4.
- Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific inquiry learning in a simulation environment. *Journal of Computer Assisted Learning*, 19, 9–20.
- Swaak, J., & De Jong, T. (1996). Measuring intuitive knowledge in science: The development of the What-if test. *Studies in Educational Evaluation*, 22, 341–362.
- . (2001a). Learner vs. system control in using online support for simulation-based inquiry learning. *Learning Environments Research*, 4, 217–241.
- . (2001b). Inquiry simulations and the assessment of intuitive knowledge. *Journal of Computer Assisted Learning*, 17, 284–294.
- Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative effects on learning. *Educational Technology Research & Development*, 53, 47 – 58.
- Tabak, I., & Reiser, B. J. (2008). Software-realized inquiry support for cultivating a disciplinary stance. *Pragmatics & Cognition Summer 2008: Special Issue Learning Technologies*.
- Thomas, R., & Hooper, E. (1991) Simulations: an opportunity we are missing. *Journal of Research on Computing in Education*, 23(4), 497–513.
- Van der Meij, H. (1990). Effects of prior knowledge on question asking. *Zeitschrift für Pädagogische Psychologie*, 4(2), 87–96.
- Van der Meij, H. (1994). Student questioning: A componential analysis. *Learning and Individual Differences*, 6, 137–161.
- Van Joolingen, W. R., & De Jong, T. (2003). Toward cost-effective production of adaptive, interactive, and intelligent educational software. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), *Authoring Tools for Advanced Technology Educational Software* (pp. 1–33). Dordrecht: Kluwer Academic.
- Van Joolingen, W. R., De Jong, T., Lazonder, A. W., Savelsbergh, E., & Manlove, S. (2005). Co-Lab: Research and development of an online learning environment for collaborative scientific inquiry learning. *Computers in Human Behavior*, 21, 671–688.
- Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: instructional design for complex learning. *Educational Psychologist*, 38(1), 5–13.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13, 21–39.
- Wilhelm, P., & Beishuizen, J. J. (2004). Asking questions during self-directed inductive learning: Effects on learning outcome and learning processes. *Interactive Learning Environments*, 12(3), 251–264.
- Xun, G. E., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research, & Development*, 52(2), 5–22.
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20, 99–149.