Engaging Students' Prior Knowledge during Instruction Improves Their Learning of Groundwater and Aquifers

Leilani A. Arthurs & Chelsie M. Kowalski

University of Colorado at Boulder

Author Note

Leilani A. Arthurs, Department of Geological Sciences, University of Colorado at Boulder; and Chelsie M. Kowalski, Department of Geological Sciences, University of Colorado at Boulder.

Correspondence concerning this article should be addressed to Leilani Arthurs, Department of Geological Sciences, University of Colorado at Boulder, Boulder, CO 80309.

Contact: Leilani.Arthurs@Colorado.EDU

ABSTRACT

Water sciences education is paramount to sustainable groundwater resource management, especially of drinking water, but misunderstandings about groundwater among non-experts remain widespread. Groundwater residence is an especially challenging concept to learn because it is not directly visible in typical circumstances. The present study uses a quasi-experimental research design to compare the impacts that two instructional sequences have on improving students' conceptual understanding of groundwater residence and aquifers. Both instructional sequences are designed to use active learning, but only one solicits and engages students' preconceptions. The theoretical framework for this study is the knowledge integration perspective of conceptual change. As such, this study considers cognitive, temporal, and social dimensions of learning. To assess students' learning, concept sketches were analyzed using diagrammatic and textual content analyses, normalized learning gains were calculated, multiple-choice items were scored dichotomously (i.e., scored as either correct or incorrect), free-response items were scored for partial credit, and classroom observations tracked social interactions. We found significantly larger learning gains when students' preconceptions were explicitly incorporated into the instructional sequence compared to when they were not taken into account. We also found the prior-knowledge instructional sequence (PKIS) positively impacted both Caucasian and non-Caucasian students as well as male and female students. Our findings indicate that actively engaging students' prior knowledge in the ways that were researched herein can be a high impact teaching practice and is worthy of future research in other specific domains beyond groundwater residence and aquifers.

Key words: Groundwater, aquifers, conceptual change, alternate conceptions, misconceptions

INTRODUCTION

Conceptions about Groundwater Residence

The term *groundwater residence* in the context of this study is used to describe where groundwater resides, is located, exists, or occurs in the subsurface in aquifers. Groundwater flows at different rates; thus, the term r*esidence time* is the length of time groundwater resides, is located, exists, or occurs in a specific subsurface reservoir or aquifer. This study is concerned with groundwater residence, not residence time.

Despite the recognized need in the US to teach students about groundwater, especially as a source of drinking water, misunderstandings and alternate conceptions about groundwater are widespread among non-experts (Agelidou et al., 2011; Arthurs & Elwonger, 2018; Pan & Liu, 2018, Unterbruner et al., 2108). An *alternate conception* is defined as an "idea or thought held … at any point in time relative to the instructional period of interest, formed by direct or inferred experience, and one that is more or less scientifically accurate and complete" (Arthurs, 2011, p. 137). Similarly, "[m]ental models are what people really have in their heads" (Norman, 1983, p. 12); they are mental representations of complex situations or problems (Derry, 1996). In this study, the terms *conception* and *mental model* are used synonymously. Mental models that are incongruent with expert-defined mental models are referred to in the literature as alternate frameworks (Diver, 1981; Dal, 2007); misconceptions (Helm, 1980); and preconceptions (Novak, 1977; Clement, 1993).

The most ubiquitous pattern of mental models about groundwater that students hold is the "separate pattern" (Arthurs & Elwonger, 2018, p. 60). This pattern of mental models is one in which students conceptualize groundwater as being separate from rock, such that water underground exists as large continuous bodies of water (e.g., underground river, underground lake, etc.). Studies reveal the separate pattern is held among 7-9th graders in Israel (Ben-zvi-Asarf & Orion, 2005); 8th graders in North Carolina, US (Dickerson & Dawkins, 2004); junior high school students in Taiwan (Pan & Liu, 2018); 7th graders and college students in Austria (Unterbruner et al., 2016); college students in Germany (Reinfried, 2006); and college students in Nebraska and Georgia, US (Arthurs & Elwonger, 2018).

Although some general education researchers and educators consider the separate pattern an inappropriate way for students to conceptualize groundwater (Dickerson et al., 2005; Unterbruner et al., 2016), water-filled underground caves and tunnels are common features in karst aquifers that form when soluble rock such as limestone dissolves (Ford & Williams, 2013; Kiraly, 2003; National Park Service, 2020; Palmer, 2007). Thus, students are not incorrect in conceptualizing groundwater residence as the separate pattern. Instead, the misunderstanding occurs in assuming that *all* groundwater resides this way. Thus, the separate pattern for conceptualizing groundwater is neither wrong nor scientifically inappropriate; however, it is scientifically incomplete because karst aquifers are only one type of aquifer. Karst landscapes cover about 20% of the U.S. (Weary & Doctor, 2014). About 40% of the US population and about 25% of the world population obtain drinking water from karst aquifers (Kalhor et al., 2019; Ghasemizadeh et al., 2012).

In the context of this study, we operationalize the term *non-karst aquifer* to mean an aquifer that does not conform to the separate pattern. For the purposes of this study on one-week-long lessons about aquifers and groundwater resources in general education introductory-level geoscience courses, it is perhaps worth noting that the level of content-specific details is not as advanced as would be for courses dedicated to hydrogeology. Instead, the level of content-specific details is constrained to introducing students to basic and typical examples of non-karst aquifers: unconfined aquifer, confined aquifer, and perched aquifer (Fetter, 2001; Freeze & Cherry, 1979; Reichard, 2011; US Geological Survey, 2020). Groundwater in these aquifers is held (a) within the fractures and pore spaces of consolidated rock or (b) between the clasts of buried and/or lithified sediments. Arthurs and Elwonger (2018) characterized the former as the 'composite internal pattern' and the latter as the 'composite external pattern' of mental models. Thus, the goal of instruction should not be dispelling the 'separate pattern' but, rather, supporting students' incorporation of the 'composite internal pattern' and the 'composite external pattern' into student mental models of groundwater residence and aquifers as sources of drinking water.

Despite decades of research describing evidence-based best practices for general instruction, the question of how to translate *general* best practices into instruction on *domain-specific* concepts, such as groundwater, remains relevant in grade levels up to and including the college level. Indeed, 'the translation process often remains elusive' (National Research Council, 2012, p. 180). Aligned with constructivist theories of teaching and learning (Ausubel & Ausubel, 2000; Driver & Erickson, 1983; Powell & Kalina, 2009), Bar (1989) and Meyer (1987) called for student-held preconceptions about groundwater to be used as teaching tools. They, however, were uncertain about *how* to use students' preconceptions as instructional tools. About three decades later, responding to the still unanswered calls, Arthurs (2019) developed a week-long instructional sequence designed to explicitly solicit and actively engage students' preconceptions about groundwater as resources for teaching about aquifers. In this study, we focus on students' prior knowledge in the form of their preconceptions. The research question driving the present study is: To what extent does this prior-knowledge instructional sequence (PKIS) aid college students in developing more expert-like ways of conceptualizing groundwater residence relative to a similar instructional sequence that does not utilize students' prior knowledge (non-PKIS)? We answer this question by addressing the cognitive, temporal, and social dimensions of students' conceptual change.

Theoretical Framework

Aligned with constructivist theories of teaching and learning (Ausubel & Ausubel, 2000; Driver & Erickson, 1983; Powell & Kalina, 2009), the present study is framed with the knowledge integration perspective of conceptual change. This perspective factors in cognitive, temporal, and social dimensions to provide a more holistic understanding of conceptual change (Linn, 2008). It also emphasizes that four practices should be a part of classroom instruction to facilitate conceptual change: (i) utilize personally relevant problems; (ii) create opportunities to make individual student thinking visible; (iii) provide students opportunities to learn from each other by sharing, discussing, and evaluating each other's ideas; and (iv) create opportunities for students to reflect on and monitor their performance (Linn, 2008).

The knowledge integration perspective suggests learning is gradual because students need time to grapple with their own confusing and conflicting ideas (Linn, 2008). Additionally, it argues the 'variability in student ideas is fundamentally a valuable feature and that instruction designed to capitalize on the variability … has [the] potential for facilitating conceptual change' (Linn, 2008; p. 715). This perspective is utilized in the present study to (i) inform the choice of the PKIS as a subject of study, (ii) inform the choice of data to collect and analyze, and (3) frame the discussion of the results.

In this study, we focus on students' prior knowledge in the form of their preconceptions. We adopt the definition of preconceptions as pre-instructional conceptions (Arthurs, 2011) that are naïve (Clement, 1993; Kinchin et al., 2000) and abstract knowledge structures associated with deep ontological commitments about how individuals make sense of the world (Vosniadou, 2014). Individuals with preconceptions are often resistant to change (Sinatra, 2005) because their preconceptions contain elements that are reinforced by individuals' experiences with the world around them (Lakoff & Johnson, 1980).

Thus, viewed through a constructivist lens, science learning cannot be achieved by the deceptively simple replacement of apparently incorrect ideas with correct ideas (Vosniadou, 2014). Instead, science learning involves confronting one's own potentially confusing and conflicting ideas (Linn, 2008) and developing a more nuanced understanding of scientific concepts in which one's preconceptions find an explicit connection to more scientifically accurate conceptions. It is with these theoretical underpinnings in mind that we strive to answer the previously stated research question.

MATERIALS AND METHODS

Methodology

This study received Institutional Review Board approval. It uses the methodology of quasi-experimental research design (Price et al., 2015). Our implementation of this methodology has nonequivalent groups, pre- and post-instruction tests, and a time-series design (Price et al., 2015). Participants in the PKIS group and the non-PKIS group were drawn from college courses in which students self-enrolled. To ensure the two groups were as similar as possible, both groups consisted of college students enrolled in introductory-level geoscience courses, had the same instructor with the same norms and expectations for class participation, and spent one week in the semester explicitly discussing aquifers. The pre- and post-instruction tests and the time-series data collected were used to investigate students' conceptions of aquifers and groundwater residence (dependent variable) at different points in time relative to each group's one-week instructional sequence on aquifers (independent variable).

Setting and Population

This research was conducted in the naturalistic setting of college classrooms. In education research, *natural setting* is defined as a realistic open situation, rather than a laboratory-based controlled situation where variables can be manipulated (Cohen et al., 2002; Green et al., 2012). The PKIS group and non-PKIS group were comprised of college students enrolled in two central USA public universities. They were enrolled in introductory-level geoscience courses that satisfy the natural science requirement for graduation. As survey courses, they addressed a range of concepts but both courses addressed groundwater residence. A timeline and list of topics covered by week in each course is provided in Supplemental Material Table S1. The population demographics for both groups are shown in Table 1.

Instructional Sequence

The courses in which the PKIS and non-PKIS were implemented are considered interactive in that the curricula were designed to actively engage students in their learning rather than only being recipients of information. Students not only listened to lectures but also participated in polls, independent work, small group work, and whole class discussions. Both curricula were developed utilizing backward design wherein learning goals are articulated and then used to inform assessments and learning activities (Wiggins & McTighe, 2011). The in-class portions of the PKIS and non-PKIS each occurred over a one-week period of instruction about aquifers and groundwater residence in which students met for three 50-minute class meetings. The PKIS was used during Week 9 of a 16-week course, and students in the PKIS group did not learn about groundwater or related concepts in the course prior to that week. The non-PKIS was used during Week 4 of a 16-week course, and students in the non-PKIS group learned about fluid storage and mobility in Week 2 of the course. At the end of the PKIS and non-PKIS, students completed a weekly homework assignment to review and apply what they learned during the week.

Prior-Knowledge Instructional Sequence (PKIS)

A single instructor designed the PKIS over six iterations of implementation over five years at two different large state universities (each with a total student enrollment greater than 30,000) in the US (Arthurs, 2019). This instructional sequence about aquifers consists of three class meetings, each 50 minutes long. All three meetings are interactive lecture periods. A detailed description of the PKIS and how to implement it can be found in Arthurs (2019). Briefly, the PKIS is comprised of a series of in-class activities that are embedded into lecture periods, thus creating interactive lectures. On Day 1 of the instructional sequence, students engage in (i) a prior knowledge check where they pair up and discuss the real-world problem of where they think the water people drink comes from, which is followed by a whole-class discussion; (ii) a prior knowledge check where they individually respond to a polling question about what they think an aquifer looks like; and (iii) an interactive video-based demonstration where they are asked to record their predictions, observations, and explanations for what happens when three drops of water are placed on four rocks with different permeabilities. On Day 2 of the instructional sequence, students engage in (i) a follow-up discussion based on the predictions, observations, and explanations they submitted on Day 1; (ii) a viewing of students' prior-knowledge drawings of groundwater and aquifers, which were collected prior to the start of the instructional week about aquifers, to explicitly show their ideas as a segue into a mini lecture about how geoscientists define three non-karst types of aquifers; and (iii) a concept sketching exercise where students are asked to apply what they learned from the mini lecture to shade in the three different types of aquifers on a provided base-form sketch. On Day 3 of the instructional sequence, students engage in an activity where representative sketches for each aquifer type collected on Day 2 are displayed for whole-class discussion about what is consistent with geoscientific conceptions of each aquifer type and what could be revised to bring a particular student sketch more closely aligned with expert-like conceptions. After this discussion, the instructor projects a blank base-form sketch and shows how water moves in the subsurface to create different aquifer types.

Non-Prior-Knowledge Instructional Sequence (non-PKIS)

A cohort of five faculty members designed the non-PKIS, and data collected during the third semester of implementation is used in the present study. This instructional sequence about aquifers consists of three class meetings, each 50 minutes long. The first two meetings are interactive lecture periods and the third meeting is a recitation period. On Day 1 of the instructional sequence, students engage in an interactive lecture where they learn about two of the non-karst types of aquifers (confined aquifers and unconfined aquifers, but not perched aquifers). Interspersed in the lecturing, students are asked two polling questions about groundwater flow and three think-pair-share questions (which geologic material makes for a good aquifer, what could perturb the water table, and identify aquitards or low permeability layers in two figures displayed on a PowerPoint slide). Students are also asked to draw the legend for two different figures of aquifers displayed on a PowerPoint slide. On Day 2 of the instructional sequence, students engage in an interactive lecture that begins with displaying accurate and clear legends that students submitted during the previous class meeting. They then answer three polling questions to review the lecture content from the previous class meeting. This is followed by three open-ended questions related to groundwater flow, in preparation to learn about how to compute groundwater discharge and to learn about hydraulic head and hydraulic gradient. Day 2 of the instructional sequence ends with students applying what they learned to calculate the hydraulic gradient for a given scenario. Day 3 of the instructional sequence is a recitation period during which students complete a worksheet about aquifers. Students complete the worksheet at their own pace and may work with peers if they like. Students retain their worksheet and complete it at home if they do not finish it during class time.

Data Sources

To examine the cognitive, temporal, and social considerations associated with the knowledge integration perspective of conceptual change, the following sources of data were used: instructor lesson plans; instructor notes about in-class activities and discussions; student responses to paper-and-pencil in-class activities; and student responses to weekly homework assignments, exams, and pre- and post-course surveys. Trained third parties made classroom observations that were also used.

The **cognitive dimension** of conceptual change that this study is interested in is conceptual learning gains. Learning gains were estimated through the analysis of students' responses to a free-response item before and after the instructional sequence. Students in the PKIS group and non-PKIS group completed an in-class activity as a gauge of their prior knowledge before the instructional sequence began. It consisted of a prompt and large blank space in which to draw and label a sketch. In this study, this type of in-class activity is called a *free-sketch activity* because students begin drawing on an entirely blank space where they are free to create a sketch from scratch. The prompts are provided in Supplemental Material Table S2. Students addressed a similar prompt in the course final exam.

The **temporal dimension** of conceptual change of interest in this study is the longitudinal development of students' mental models akin to karst and non-karst aquifers, at four different times during the semester. The prompts are provided in Supplemental Material Table S2. A timeline of the sampling points relative to the weeks in the semester is provided in Supplemental Material Table S1.

For the PKIS group, the first time point (T_1) was the week before the instructional sequence began (Week 8 of the semester). The second time point (T_2) was during the instructional sequence in Week 9, after a mini lecture that separately described three main types of aquifers. The third time point (T_3) was three weeks after the instructional sequence ended (during a mid-term exam in Week 11 of the semester, after a week of vacation). The fourth and last time point (T_4) was five weeks after the instructional sequence ended (during the final exam in week 16 of the semester). In contrast to the free-sketch activities at T_1 and T_4, the assessments at T_2 and T_3 are *delimited-sketch activities*. A delimited-sketch activity is an in-class activity that begins with a partial sketch already provided, which is called a *base-form sketch* (Arthurs, 2019). Students add to the base-form sketch by drawing additional features, amending existing features, and labeling their sketch to help clarify their ideas. The delimited-sketch activity at T_2 is an activity in the PKIS.

For the non-PKIS group, the first time point (T_1) was two weeks before the instructional sequence began (Week 2 of the semester). The second time point (T_2) was during the instructional sequence in Week 4, after a lecture describing two types of aquifers. The third time point (T_3) was two weeks after the instructional sequence ended (during a mid-term exam in Week 6 of the semester). The fourth and last time point (T_4) was 11 weeks after the instructional sequence ended (during the final exam in week 16 of the semester). The non-PKIS dis not utilize a delimited-sketch activity, and students completed free-sketch activities at all four time points.

The in-class activities for the PKIS group at T_1 and T_2 were formative assessments used to draw out student thinking for explicit discussion in subsequent lessons. As with other in-class activities during the semester, the instructor let students know their responses would aid her in better understanding their current thinking, which would then assist her in helping them take what they know to the next level. By 'current thinking,' the instructor meant students' current ideas about groundwater residence. By 'the next level,' the instructor meant advancing students' ideas to achieve course learning goals associated with the course. Students earned two participation points toward the in-class activity component of the course for completing their in-class activities clearly and demonstrating good-faith effort at communicating their ideas, not for correctness. As formative assessments, de-identified responses were displayed on PowerPoint slides and discussed in subsequent lessons in the instructional sequence to summarize the diversity of ideas expressed and to build on those ideas through follow-up in-class activities and discussions. Although several examples of representative student work were displayed and discussed during class, in-class activities were not returned to students.

The in-class activity soliciting students' ideas about groundwater residence for the non-PKIS group at T_1 and T_2 were not formative assessments and were not discussed in subsequent lessons. They were intended only to obtain insights about students' mental models for this study. Nevertheless, as with the PKIS group, students in the non-PKIS group were also informed they would earn two participation points toward the in-class component of the course for completing the in-class activities clearly and demonstrating good-faith effort at communicating their ideas, not for correctness. As with the PKIS group, in-class activities were not returned to students in the non-PKIS group.

The assessments at T_3 and T_4 for the PKIS group and non-PKIS group were summative assessments. As summative assessments, the results of these exams were not discussed as part of any follow-up instruction in class. Responses to other non-sketch-based homework and exam items as well as the instructor's notes about in-class discussions and interactions with the students provide additional information about the longitudinal development of students' mental models.

The **social dimension** of conceptual change of interest in this study are mainly students' interactions during the focal instructional sequence. To investigate these interactions, sources of data include: the instructor's lesson plans and notes about student interactions during in-class activities and discussions, student responses to pre- and post-course survey items (Supplemental Material Table S3), and classroom observations. Two trained observers external to the course and department used the Classroom Observation Protocol for Undergraduate Science (COPUS) (Smith et al., 2013) to observe the PKIS, and one faculty peer reviewer and one graduate student teaching assistant recorded in-class observations in the non-PKIS course. The two trained observers were staff members who were employed and trained by a campus office that makes formal observations of classroom teaching.

Data Analysis

For both the PKIS group and non-PKIS group, time points T_1, T_2, and T_3 were similarly spaced relative to one another, which allows for more direct comparisons between the two groups. Time point T_4 in the PKIS group occurs five weeks after T_3, and T_4 in the non-PKIS group occurs 11 weeks after T_3. Thus, the pre- and post-instruction tests for the purposes of measuring short-term conceptual learning gains in this study occur at T_1 and T_3. Data collected at T_4 are used to make supplemental observations about longer-term conceptual learning gains (see *Temporal Dimension* in the Results section).

Concept Sketch Analyses

Sketches created during free- and delimited-sketch activities were analyzed using diagrammatic (Gobert, 2000) and textual content analysis (Sapsford, 1999). A double-coding process (Krefting, 1991) was applied to all pre-instructional annotated sketches, with two weeks between the first and second coding session, to determine the types of mental models displayed using an author-developed rubric. The rubric was used to classify them as underground pockets, caves, caverns; pools, lakes; reservoirs; rivers, layers of water, tunnels; and pipes and veins (Supplemental Material Figure S1). With greater than 97% agreement between the two coding iterations, the process yielded little to no discrepancies in the codes. The high percent agreement provides support of reliability (Krefting, 1991).

Analyses of all annotated sketches were performed with a scoring rubric (Supplemental Material Table S4). The authors developed the scoring rubric to evaluate specific features in concept sketches against an expert standard. The expert standard used for this study is based on the descriptions of perched aquifers, unconfined aquifers, and confined aquifers provided in two respected discipline-specific textbooks, one on hydrogeology (Fetter, 2001) and one on groundwater (Freeze & Cherry, 1979) as well as the introductory-level course textbook (Reichard, 2011). Using these descriptions, one researcher with a geology and education background developed a scoring rubric in which the highest score that could be assigned to a concept sketch is '6' and these scores are then translated into percentages (6 points = 100%). The closer to 100% a concept sketch scores, the more expert-like the communicated mental model about groundwater residence is. After the scoring rubric was developed, two research assistants with a geology background critiqued the rubric for content, clarity, and organization. All three researchers discussed the critiques and no changes to the rubric were deemed necessary as it was clear and consistently applied. The critiques contributed to the rubric's trustworthiness (Guba, 1990).

Two researchers (a professor and a professional research assistant) independently coded all the sketches and then compared their coding results to achieve interrater agreement (LeBreton & Senter, 2008). A comparison of scores assigned to sketches from the PKIS group resulted in >84% initial interrater agreement, and after discussion resulted in 100% interrater agreement. A comparison of scores assigned to sketches from the non-PKIS group resulted in >84% interrater agreement prior to discussion. Discussion of the sketches from the non-PKIS group also led to adjustments in the coding rubric to accommodate sketched features not previously observed. For example, some non-PKIS post-instruction sketches included a water table as a layer with depth or thickness rather than as a boundary between the unsaturated and saturated zones. Application of the adjusted coding rubric resulted in interrater agreement >88%, and after discussion resulted in 100% interrater agreement. Although the PKIS curriculum addressed all three types of non-karst aquifers, the non-PKIS curriculum did not explicitly address perched aquifers. The reason for this difference is attributed to curriculum design priorities of those involved in designing these curricula. Thus, comparative statistical analyses were performed using data for only confined and unconfined aquifers collected from the PKIS group and non-PKIS group. The PKIS group had two free-sketch activities (T_1 and T_4) and two delimited sketch activities (T_2 and T_3) while the non-PKIS group had free-sketch activities at all four time points (T_1, T_2, T_3, and T_4) because the delimited sketch activities were part of the PKIS curriculum (with test group) and not part of the non-PKIS curriculum (with control group).

Data collected at T_1, T_2, T_3, and T_4 were statistically analyzed (i.e., calculated average, standard error, t-value, p value, and Cohen's d) to examine the overall development of students' mental models over time. Additionally, each group's average learning gains were computed using Equation 1 (Hake, 1998). In Equation 1, <g> is called the normalized gain, <pre> is the group's average pre-instruction test score, <post> is the group's average post-instruction test score, and the denominator equals the maximum possible gain. Learning gain, <g>, is reported as a fraction of 1, where 1 represents 100%.

(1) $$< g >= \frac{<post> - <pre>}{100 - <pre>}$$

Potentially statistically significant differences from one time to another were determined using a two-tailed t-test, and the effect size was determined by calculating Cohen's d. The same analyses were performed to determine potential impacts based on gender and race. To determine the impact of the PKIS curriculum and non-PKIS curricula over time, learning gains were calculated with data of students for whom data was collected at all four time points. For the analyses of learning gains, the PKIS group consists of 51 students and the non-PKIS group consists of 52 students. This is about 84% and 94% of the participants in the PKIS group and non-PKIS group, respectively, and they are demographically representative of their groups.

Homework and Exam Items

Additional insights into the development of students' mental models about groundwater were obtained via answers to groundwater-related items in homework and exams. Although the homework and exam items in the PKIS and non-PKIS courses were not directly comparable because they are not the same, they did provide an additional means to gauge student understanding. These items were mainly multiple-choice items. A few were open-ended items. Multiple-choice items were scored dichotomously (i.e., scored as either correct or incorrect), and free-response items were scored to permit partial credit for partly correct answers. These items are included in Supplemental Material Tables S5 and S6.

Classroom Observations

Two trained observers used the COPUS to observe all three days of the PKIS, and they produced a report describing their observations. A faculty peer observer and a graduate teaching assistant observed the non-PKIS course, and they provided reports describing their observations. The graduate teaching assistant observed all three days of the non-PKIS and the faculty peer observer was present only one day. Information in these reports were used in conjunction with the instructor's lesson plans and notes to aid in characterizing the social dimensions of conceptual change.

Pre- and Post-Course Surveys

Additional insights into the social dimensions of conceptual change were obtained via responses to items on the pre- and post-course surveys. Matching Likert-scale items were analyzed to determine whether any pre/post shifts in students' attitudes towards working alone and with others might have occurred. Two free-response items in the post-course survey were analyzed to determine the frequency with which social aspects of the course that students mentioned. Social aspects counted included in-class activities, discussions, and group work.

RESULTS

The Results section addresses the three dimensions of learning presented in the theoretical framework: cognitive, temporal, and social.

Cognitive Dimension

The depth of conceptual understanding was determined by comparing rubric scores for the concept sketches that the PKIS group and non-PKIS group drew at T_1 and T_3, where T_1

represents pre-instruction conceptual understanding and T_3 represents post-instruction conceptual understanding. The results show the PKIS group and non-PKIS group had pre-instructional conceptual understandings that were similar (Figure 1 and Table 2a). Representative examples of students' pre-instructional concept sketches are shown in Figure 2. The results show a large and statistically significant positive shift towards more expert-like mental models for the PKIS group from T_1 to T_3 (Table 2b). The results also show a statistically significant positive shift towards more expert-like mental models for the non-PKIS group from T_1 to T_3 (Table 2c). Additionally, the results show the PKIS group and non-PKIS group had post-instructional conceptual understandings that are significantly different (Table 2d). Finally, the results show a statistically significant difference in the overall learning gains between the PKIS group and the non-PKIS group (Table 2e).

Temporal Dimension

Students' mental models about groundwater residence akin to karst and non-karst aquifers were compared among the PKIS group and non-PKIS group at four different time points.

Temporal Change in PKIS Group

For the PKIS group, the most commonly occurring pre-instructional mental model of groundwater residence at T_1 was that it exists in underground 'rivers' or layers of water (26%). The second most common was underground 'pools,' 'lakes,' or 'reservoirs' of water (23%). Twenty-three percent also expressed water is intermixed with soil. The third most common are groundwater resides in underground 'pockets,' 'caves,' or 'caverns' (21%). Only 7% expressed groundwater resides in the spaces in between small rocks, gravel, and sand. Also, only 5% expressed groundwater resides inside porous or permeable rock itself. The results reveal a gradual change toward more expert-like mental models over time. Evidence of students' conceptual change over time were determined by applying the scoring rubric to concept sketches collected at T_1, T_2, T_3, and T_4 (Figure 3a and Figure 4a). Additional evidence of a change toward more expert-like ways of conceptualizing groundwater residence comes from students' performance on groundwater-related items in homework and exams, which have a combined average of >80% (see Supplemental Material Table S5). The PKIS curriculum had a large positive impact on the conceptual development of male and female students as well as Caucasian and non-Caucasian students (see Supplemental Material Tables S7).

Temporal Change in non-PKIS Group

For the non-PKIS group, the most commonly occurring pre-instructional mental model of groundwater residence at T_1 was it resides in underground 'pools,' 'lakes,' or 'reservoirs (36%). The second most common was underground 'pockets,' 'caves,' or 'caverns' of water (35%). The third most common was groundwater is found in underground 'rivers' or layers of water (24%). Only 4% expressed groundwater resides in the spaces in between small rocks, gravel, and sand. Also, only 2% expressed groundwater resides inside porous or permeable rock itself. The results reveal a gradual change toward more expert-like mental models over time. Evidence of students' conceptual change over time were determined by applying the scoring rubric to concept sketches collected at T_1, T_2, T_3, and T_4 (Figure 3b and Figure 4b). Additional evidence of a change toward more expert-like ways of conceptualizing groundwater residence comes from students' performance on groundwater-related items in homework and exams, which have a combined average of >80% (see Supplemental Material Tables S6). The non-PKIS curriculum had a positive impact on the conceptual development of male and female students as well as Caucasian and non-Caucasian students (see Supplemental Material Table S8).

Social Dimension

Social Interactions in PKIS Group

The instructor's lesson plans, instructor's notes, and external observers' reports show the PKIS is best described as an interactive lecture with a conversational tone characterized by the back-and-forth sharing of ideas between students and between students and the instructor. Each class meeting was facilitated with 19–20 PowerPoint slides. Of them, 10% were used as visual transitions from one topic to another and/or as announcements, 15% engaged students as part of lecture, 35% were used to transfer information via lecture, and 40% were used to facilitate in-class activities and discussions. During times of lecture, students were actively engaged in note taking, listening, and asking questions. During times of individual work, students engaged in independent thought and committed their ideas to paper. During group work and whole-class discussion, the room was vibrant with audible discussion, inquiry, and even laughter.

An analysis of the prompts shown in Supplemental Material Table S3, reveals the pre- and post-course survey results show there were no shifts in the extent to which the PKIS group liked working alone, working with other people, and their preferences for working in one way or another. A free-response item on the post-course survey shows 3% of students wanted 'fewer in-class activities' or 'less participation.' Meanwhile 52% of students said they enjoyed one or more social aspects of the course. Of these students, six (12%) offered suggestions for doing more socially oriented activities. Representative quotes that highlight these sentiments among the PKIS group are listed below.

- "I loved how [the interactive] lectures were always fun and interesting. I often became involved learning about subjects that I had little interest in to begin with. I really enjoyed taking this class!"

- "I really enjoyed the [the interactive] lectures, and the knowledge [the professor] had on the subject. Some of the material can be dry, and she made it fun with activities and videos."
- "I really enjoyed the professor and the people in the class! I learned a lot from the in class activities as well."
- "I really liked the open discussion in class. I also enjoyed the additional videos that we watched. Also, I really did enjoy [the professor] as a teacher and appreciated her enthusiasm and encouragement [to participate in class]."
- "What I liked best was the encouragement to discuss and question ideas we talked about both with our classmates individually and as a whole class."
- "I loved how [the professor] incorporated the class's ideas into the next powerpoint. She was very engaging and made me want to learn. This class was never a chore and was always fun."
- "Include more group presentations and models."
- "I feel the instructor can use small group work more often …."

Social Interactions in non-PKIS Group

The instructor's lesson plans, instructor's notes, faculty peer observer's report, and graduate student teaching assistant's observation report show the non-PKIS can be described as an interactive lecture-and-recitation. The lecture periods have a conversational tone characterized by the back-and-forth sharing of ideas between students and between students and the instructor. Each lecture period was facilitated with 17 PowerPoint slides (Day 1) or 15 PowerPoint slides (Day 2). Of them, 16% were used as visual transitions from one topic to another and/or as announcements, 50% engaged students as part of lecture, 47% were used to transfer information via lecture, and 6% were used to facilitate in-class activities and discussions. The total percentages do not sum to 100% because some slides both engaged students as part of lecture (e.g., with polling question) and transferred information via lecture. During times of lecture, students in the non-PKIS were actively engaged in note taking, listening, and asking questions. During times of individual work, they engaged in independent thought and committed their ideas to paper. During group work and whole-class discussion, the non-PKIS group exhibited audible discussion and inquiry.

The pre- and post-course survey results show there were no shifts in the extent to which the non-PKIS group liked working alone versus working with other people. A chi-square test of independence was performed to examine the relation between pre/post non-PKIS curriculum (independent variable) and the preference for working alone and/or with others (dependent variable). The relation between these variables was statistically significant, X^2 (3, N = 49) = 0.295, p = .990. By the end of the course, students in the non-PKIS group were more likely to prefer a combination of both working alone and with others. A free-response item on the post-course survey shows 2% of students wanted 'less group work'. Meanwhile 29% of students said they enjoyed one or more social aspects of the course. Representative quotes that highlight these sentiments among the non-PKIS group are listed below.

- "I liked [the professor's enthusiasm. It really helped me get in the mindset to learn and to enjoy the class discussion."
- "The topic is very interesting so I enjoyed spending time learning it and the people made the course better."
- "The fact that [the professor] made an effort to learn everyone's names in such a large class and actually wanted to talk to you personally and learn about you really showed

how much she cares about her students. Seeing her commitment to her students made me even more motivated to put effort into the class and do well."

DISCUSSION

The Discussion section addresses the three dimensions of learning presented in the theoretical framework: cognitive, temporal, and social.

Cognitive Implications

The PKIS group and non-PKIS group had similar levels of conceptual understanding about groundwater residence and aquifers as a drinking water source at T_1, prior to their respective instructional sequences about aquifers and groundwater residence (Table 2, Figure 1, and Figure 2). Figure 2 highlights examples of the ways in which students sketched continuous bodies of water versus water in the interstices of sediment. Both groups held similar preconceptions about continuous underground bodies of water that also appear across grade levels and regions (Arthurs & Elwonger, 2018; Ben-zvi-Asarf & Orion, 2005; Dickerson & Dawkins, 2004; Pan & Liu, 2018; Reinfried, 2006; Unterbruner et al., 2016).

Although the majority in both groups held mental models of continuous underground bodies of water akin to karst aquifers, there were two differences in the spread of mental model types at T_1 between the PKIS group and non-PKIS group. The first difference is that about 7% of students in the non-PKIS group indicated groundwater we use for drinking water comes from 'pipes' or 'veins' and the PKIS group students made no mention of pipes or veins. Based on previous research (Arthurs & Elwonger, 2018; Dickerson & Dawkins, 2004), this is a far less commonly held conception than others noted in the literature. Thus, it is perhaps not surprising that a few students in the non-PKIS group expressed this idea and none in the PKIS group did.

The second difference is that about 23% of PKIS group students indicated groundwater we use for drinking water is intermixed with soil and the non-PKIS group students made no mention of water intermixed with soil. Given that water intermixed with soil is not a commonly identified preconception in the research literature (Arthurs & Elwonger, 2018) and given that the PKIS group's T_1 prior knowledge check occurred at the end of a week-long lesson about soil resources, we hypothesize that this mental model appeared in the PKIS group because they learned about soil resources in the previous week of instruction. Although soil holds water moisture, soil is neither an aquifer nor is soil moisture a source of drinking water. In this sense, the PKIS group started their instructional sequence about aquifers and groundwater residence with a misconception that the non-PKIS group did not express.

Despite these two differences, the PKIS group and non-PKIS group began their respective instructional sequences about aquifers and groundwater with comparable levels of conceptual understanding and similar types of preconceptions relevant to karst and non-karst aquifers. One reason why the conception of large continuous bodies of water underground may be so common among students in this study and others is that the conceptions of 'underground lake' and 'underground river' serve as metaphorical tools. Lakoff and Johnson (1980) suggest such metaphors enable people to use what they know based on their direct physical experiences to understand more abstract or not directly visible phenomena. In other words, applied to this context, students use their direct experiences seeing or recreating in lakes and rivers to understand the typically unseen underground environment of aquifers and groundwater.

Comparisons of the average learning gains computed using the pre- and post-instruction results (at T_1 and T_3) reveal that the PKIS and the non-PKIS approaches to teaching and learning about aquifers and groundwater residence both facilitated a shift towards more expert-like conceptions. Both the PKIS group and non-PKIS group exhibited statistically significant learning gains from T_1 (pre-instruction) to T_3 (post instruction) that were more than twice that expected from traditional lecture-based instruction, according to research by Hake (1998) who found lecture alone leads to about 0.23 learning gains at most and Freeman et al. (2014) who conducted a meta-analysis demonstrating active learning leads to greater learning than lecture alone. See Table 2e for the learning gains data. A plausible explanation for the difference in learning gains with these two approaches relative to traditional lecture-based instruction alone is that both the PKIS and non-PKIS curricula were designed to be interactive and utilized active learning techniques.

While the PKIS and non-PKIS curricula produced higher learning gains than expected in traditional lecture-based instruction, the PKIS curricula also produced significantly larger learning gains from T_1 (pre-instruction) to T_3 (post-instruction) compared to the non-PKIS curricula. One might argue the difference is due to differences in how the T_3 data were collected because the PKIS group used a delimited sketch activity and the non-PKIS group used a free-form sketch activity. However, the idea that the observed difference is due to the testing mode is weakened by the fact that the PKIS group used a free-form sketch activity at T_4 with very similar results as at T_3. The main discernable difference between the two instructional sequences is not whether they were interactive but whether they explicitly solicited and actively engaged students' prior knowledge to facilitate the development of more expert-like conceptions of aquifers and groundwater. Thus, it is plausible that the difference in observed learning gains between the PKIS group and the non-PKIS group is attributable to that difference. Also, recall the knowledge integration perspective of conceptual change advocates four practices in classroom instruction: (1) using personally relevant problems; (ii) making individual student thinking visible; (iii) enabling students to learn from one another by sharing, discussing, and evaluating one another's ideas; and (iv) providing students with opportunities to reflect on and monitor their performance. Although the PKIS and non-PKIS curricula addressed the personally relevant question of where people obtain their drinking water, only the PKIS curriculum implemented the other three practices by focusing on students' prior knowledge and evolving ideas.

Since the time of Meyer's (1987) and Bar's (1989) calls to utilize students' preconceptions about groundwater as instructional tools, a review of the literature reveals those calls have not been taken up until recently (Arthurs, 2019). Unterbruner et al. (2016) utilized student preconceptions documented in the literature to develop a multimedia learning program that students navigated through on their own, but they did not actively elicit and incorporate individual students' preconceptions into the learning program. Their decision was made from concern based on Sinatra's work (2005) that acknowledging students' preconceptions in a statistically significant way would reinforce misconceptions. To the best of our knowledge, we are the first to demonstrate the efficacy of explicitly invoking and directly utilizing students' preconceptions in learning about groundwater residence and aquifers.

Temporal Implications

The PKIS group and the non-PKIS group held a similar range of pre-instructional conceptions about groundwater residence, and the instructional sequence both groups experienced had a positive impact on facilitating conceptual change. The PKIS group experienced significantly higher learning gains compared to the non-PKIS group from T_1 to T_3. Although a one-to-one comparison of the longer-term retention (i.e., at T_4) of more expert-like conceptions about aquifers and groundwater residence between the PKIS group and the non-

PKIS is not possible because of a three-week difference between T_4 for the two groups, it is nevertheless possible to make within-group comparisons with the available data.

For the non-PKIS group, there was a statistically significant difference between the scores at T_3 and T_4, indicating a loss in conceptual gains 13 weeks after the instructional sequence (Table 2 and Figure 3b). This finding was not surprising because it is common for students to forget what they learned, especially when what they learned is not used regularly (Wixted, 2005). Interestingly, that loss brought the PKIS group back to the T_2 level of conceptual understanding, which is still significantly greater than their conceptual understanding at T_1, indicating statistically significant memory retention.

For the PKIS group, there was no statistically significant difference between the scores at T_3 and T_4, indicating the more expert-like conceptual understanding of aquifers and groundwater residence persisted for a large fraction of students even 8 weeks beyond the instructional sequence (Table 2 and Figure 3a). This finding was surprising because students often do forget what they learned (Murre & Dros, 2015; Terada, 2017). The overall learning gains for the PKIS group remaining relatively steady from T_3 and T_4 suggests that explicitly soliciting and actively incorporating their preconceptions into the instructional sequence had a relatively long-lasting impact. Whether or not that impact would have held even longer is an area of potential future research.

The longitudinal results for both the PKIS group and non-PKIS group from T_1 to T_4 (Figure 3) indicate conceptual change is occurs at various rates for different students and is, generally, a more gradual process rather than a rapid or revolutionary process. This is consistent with Linn's (2008) knowledge integration perspective of conceptual change. Additionally, students struggle to assimilate and/or accommodate scientific conceptions into their preconceptions. For example, Figure 4b illustrates how one student modified their initial mental model of groundwater residence after learning about permeable and impermeable rocks by adding a casing of impermeable rock around the pocket or capsule of water underground while retaining the notion that water is stored in such pockets. These empirical data support the knowledge integration perspective, which posits learning is gradual (Linn, 2008) because students need time to confront their own perhaps confusing and conflicting ideas (Linn, 2008). During the instructional period (e.g., lesson, instructional sequence, semester, etc.), time for active learning and reflection during purposeful in-class activities are required for students to make sense of new ideas learned in class and to reconcile them with their pre-existing ideas, which may be incongruent or dissonant with new ones.

Social Implications

The COPUS results and reports that the third-party reviewers provided corroborate the opportunities described in the instructor's lesson plans and notes on the PKIS and non-PKIS curricula and their implementation. In line with the knowledge integration perspective of conceptual change, the in-class activities in the PKIS allowed students to (i) individually engage in a personally relevant issue (i.e., groundwater as a drinking water source); (ii) make their individual thinking visible to themselves, their peers, and instructor; (iii) learn from one another by sharing, discussing, and evaluating one another's ideas; and (iv) reflect on and monitoring their performance. While the non-PKIS curriculum also engaged students in the same personally relevant issue of drinking water resources, it did not incorporate the other three practices.

Although not specific to the instructional sequences, the results of two free-response items in the post-course surveys indicate that many students in both the PKIS group and non-PKIS group valued various social aspects of the courses and their learning experiences in it.

These responses provide some affective insights into how students value social interactions as a part of their learning experiences, including the importance of an instructor learning their names. The courses had no statistically significant impact on how members of the PKIS group and the non-PKIS group liked working individually and with others. Both the PKIS and non-PKIS curricula positively impacted students who preferred to work alone, those who preferred to work with others, and those who had no preference.

Limitations

This study is limited in its use of (a) concept sketches as the mechanism for obtaining insights into students' mental models, (b) naturalistic settings, and (c) quasi-experimental research design. Concept sketches elicited two different but complementary ways of communicating students' ideas, diagrammatic and textual communication. Although a concept sketch can communicate key elements of an individual's mental model, it does not necessarily communicate all elements that may be present (Clement, 1982; Henriques, 2002; Osborne & Wittrock, 1983). In addition, sketching to communicate one's ideas has similar goals and limitations as verbal communication – the goal of clarity in the conveyance of ideas and the potential for imperfect conveyance of those ideas. Despite its limitations, sketching has a long and demonstrated history as a useful tool for studying mental models and cognitive development (e.g., Piaget, 1956; Rees, 2018; Roberts & Russell, 1975).

Research conducted in naturalistic settings is less controlled than in research laboratory settings, with greater opportunities for confounding variables to be introduced. For example, readers may find a drawback to the study is that the course in which the PKIS was implemented addressed three aquifer types and the course with the non-PKIS addressed only two. However, systematic comparisons were still possible given the PKIS and non-PKIS addressed two of the same aquifer types. Additionally, not all students completed all assessments related to this study due to class absences. Although research in naturalistic settings is less controlled than research laboratory settings, naturalistic settings offer opportunities to investigate learning phenomena in the actual settings in which learning occurs. While this contributes to lower internal validity, it contributes to greater external validity (Price et al., 2015). In this way, as Barab (2006) notes, learning sciences research recognizes ‘context is not simply a container within which the disembodied “regularities” under study occur, but is an integral part of the complex causal mechanisms that give rise to the phenomenon under study (Maxwell, 2004).’

Although students were not randomly assigned to the PIKIS and non-PKIS groups, these groups were representative of the courses from which they were derived and the larger population of undergraduate students at each institution at the time this research was conducted (Table 1). Furthermore, the two groups were similar in terms of demographics and their baseline conceptual understanding of aquifers and groundwater residence (Table 1, Table 2, Figure 1, and Figure 3). Using a quasi-experimental research design in two different naturalistic settings enhances the findings’ generalizability and their relevance for curriculum design and instructional decisions in groundwater-related instruction.

CONCLUSION

The idea of utilizing students’ prior knowledge during instruction has deep roots in education theory as a form of best practice (National Research Council, 2000); however, translating this general best practice to domain-specific instruction has remained elusive (National Research Council, 2012). Actively acknowledging students’ prior knowledge may seem undesirable to instructors who view them as potential barriers to learning (Sinatra, 2005; Unterbruner et al., 2016) and who believe they must be replaced with correct ideas (Bransford, 2000; Meyer, 2004). This study's findings, however, provide evidence to support Meyer's (1987) and Bar's (1989) hypothesis that students' prior knowledge can be effective instructional tools for teaching students about groundwater residence and aquifers. Major findings of this study include:

- The interactive instructional sequences for both the PKIS group and non-PKIS group produced statistically significant post-instruction learning gains, far greater than that expected with traditional lecture alone.
- The interactive instructional sequence that explicitly solicited and actively engaged students' preconceptions (PKIS group) resulted in learning gains that were significantly larger compared to the instructional sequence that did not (non-PKIS group).
- Active engagement of students' preconceptions about groundwater can lead to statistically significant learning gains for Caucasian and non-Caucasian students and male and female students.
- The results of the PKIS and non-PKIS curricula reveal students' trajectory toward more expert-like conceptual understanding varies for different students, and there remain variations in individual conceptual understanding at each time point in the trajectory. In other words, the development of different individuals' conceptual understanding does not necessarily progress at the same rate given the same instructional interventions.

These findings suggest the positive impact on learning that interactive engagement has can be increased through the explicit solicitation and active engagement of students' preconceptions. We found that the explicit solicitation and active engagement of students' preconceptions comparably benefitted male and female students as well as Caucasian and non-Caucasian students, which suggests this instructional approach can be used to support racial and gender equity and inclusion in the geosciences and STEM more broadly to some extent. For example, the PKIS has not yet been modified for visually impaired students. Additionally, our findings demonstrate that statistically significant learning gains in students' conceptual understanding of aquifers and groundwater residence can be achieved with basic instructional tools (e.g., PowerPoint and handouts) and do not require field trips, special software, or specialized apparatus. The fact that the PKIS can be implemented at low cost with standard instructional resources means that the PKIS is also accessible to many instructors who might like to include it in their courses. Additionally, the PKIS could be used as an introduction to lessons about specifically named local aquifers and groundwater resources. Finally, these findings are consistent with the knowledge integration perspective of conceptual change, which posits 'variability in student ideas is fundamentally a valuable feature and that instruction designed to capitalize on the variability … has [the] potential for facilitating conceptual change' (Linn, 2008, p. 715).

Aligned with the knowledge integration perspective of conceptual change, the PKIS uses the repertoire of characterized student-held ideas documented in Arthurs & Elwonger (2018) as 'a way to increase the efficiency and effectiveness of instruction' (Linn, 200, p. 716). The widely-held commonalities in pre-instructional conceptions about groundwater residence across grade levels and geographic regions (e.g., Arthurs & Elwonger, 2018; Ben-zvi-Asarf & Orion, 2005; Dickerson & Dawkins, 2004; Pan & Liu, 2018; Reinfried, 2006; Unterbruner et al., 2016) imply that the PKIS curriculum has applicability and the potential to positively impact the learning of student groups beyond those in this study.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1821709. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

L. Arthurs, (2011), What college-level students think: Student cognitive models of geoscience concepts, their alternate conceptions, and their learning difficulties. In A. Feig and A. Stokes (Eds.), *GSA Special Papers: Qualitative Inquiry in Geoscience Education Research* (pp. 135-152). Boulder: Geological Society of America.

Arthurs, L. A. (2019). Using student conceptions about groundwater as resources for teaching about aquifers. *Journal of Geoscience Education*, *67*(2), 161-173.

Arthurs, L. A., & Elwonger, J. M. (2018). Mental models of groundwater residence: A deeper understanding of students' preconceptions as a resource for teaching and learning about groundwater and aquifers. *Journal of Astronomy & Earth Sciences Education*, *5*(1), 53-66.

Agelidou, E., Balafoutas, G., & Gialamas, V. (2001). Interpreting how third grade junior high school students represent water. *International Journal Education Information, 20*: 19-36.

Ausubel, D. P., & Ausubel, D. P. (2000). *The acquisition and retention of knowledge: A cognitive view*. Dordrecht: Kluwer Academic Publishers.

Bar, V. (1989). Children's views about the water cycle. *Science Education*, 73(4), 481-500.

Barab, S. (2006). Design-based research: A methodological toolkit for the learning scientist. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 153-169). New York, NY: Cambridge University Press.

Ben-zvi-Assarf, O., & Orion, N. (2005). A study of junior high students' perceptions of the water cycle. *Journal of Geoscience Education*, *53*(4), 366-373.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press.

Clement, J. (1982). Students' preconceptions in elementary mechanics. *American Journal of Physics*, 50, 66-71.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics: *Journal of Research in Science Teaching*, *30*(10), 1241-1257.

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. New York: Routledge.

Dal, B. (2007). How do we help students build beliefs that allow them to avoid critical learning barriers and develop a deep understanding of geology? *Eurasia Journal of Mathematics, Science & Technology Education*, *3*(4), 251-269.

Derry, S. (1996). Cognitive schema theory in constructivist debate. *Educational Psychologist*, *31*(3/4), 163-174.

Dickerson, D., & Dawkins, K. (2004). Eighth grade students' understandings of groundwater. *Journal of Geoscience Education*, *52*(2), 178-181.

Dickerson, D. L., Penick, J. E., Dawkins, K. R., & Van Sickle, M. (2007). Groundwater in science education. *Journal of Science Teacher Education*, *18*(1), 45-61.

Driver, R. (1981). Pupils' alternative frameworks in science. *European Journal of Science Education*, *3*(1), 93-101.

Driver, R., & Erickson, G. (1983). Theories-in-action: Some theoretical and empirical issues in the study of students' conceptual frameworks in science. *Studies in Science Education*, *10*(1), 37-60.

Fetter Jr., C. W. (2001). *Applied Hydrogeology*. Prentice-Hall Inc.

Ford, D., & Williams, P. D. (2013). *Karst hydrogeology and geomorphology*. Southern Gate, West Sussex: John Wiley & Sons.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

Freeze, R.A., & Cherry, J.A. (1979). Groundwater. Prentice-Hall Inc.

Ghasemizadeh, R., Hellweger, F., Butscher, C., Padilla, I., Vesper, D., Field, M., & Alshawabkeh, A. (2012). Groundwater flow and transport modeling of karst aquifers, with particular reference to the North Coast Limestone aquifer system of Puerto Rico. *Hydrogeology Journal*, *20*(8), 1441-1461.

Gobert, J. D. (2000). A typology of causal models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education*, *22*, 937-977.

Green, J. L., Camilli, G., & Elmore, P. B. (Eds.). (2012). *Handbook of complementary methods in education research*. Washington DC: Lawrence Erlbaum Associates, Publishers.

Guba, E. G. (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 17–30). Newbury Park, CA: Sage.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, *66*(1), 64-74.

Helm, H. (1980). Misconceptions in physics amongst South African students. *Physics Education*, *15*(2), 92.

Henriques, L. (2002). Children’s ideas about weather: A review of the literature. *School Science and Mathematics*, *102*(5), 202-215.

Kalhor, K., Ghasemizadeh, R., Rajic, L., & Alshawabkeh, A. (2019). Assessment of groundwater quality and remediation in karst aquifers: A review. *Groundwater for sustainable development*, *8*, 104-121.

Kinchin, l. M ., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development: *Educational Research*. 12(1), 13-57.

Kiraly, L. (2003). Karstification and groundwater flow. *Speleogenesis and evolution of karst aquifers*, *1*(3), 155-192.

Koomson, C. K., & Owusu-Fordjour, C. (2018). Misconceptions of senior high school science students on evaporation and water cycle. *European Journal of Research and Reflection in Educational Sciences Vol*, *6*(5), 13-28.

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, *45*(3), 214-222.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods*, *11*(4), 815-852.

Linn, M. C. (2008). Teaching for conceptual change: Distinguish or extinguish ideas. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 694-722). New York, NY: Routelage.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, *33*(2), 3-11.

Meyer, H. (2004). Novice and expert teachers' conceptions of learners' prior knowledge. *Science education*, *88*(6), 970-983.

Meyer, W. B. (1987). Vernacular American theories of earth science. *Journal of Geological Education*, *35*, 193-196.

Murre, J. M., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PloS one*, *10*(7), e0120644.

National Park Service. (2020, August 16). *Karst landscapes*. Retrieved from https://www.nps.gov/subjects/caves/karst-landscapes.htm

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: National Academies Press.

National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

Norman, D. A. (1983). *Some observations on mental models.* In D. Gentner & A. L. Stevens (Eds.), Mental Models, pp. 7-14. Hillsdale, NJ: Laurence Erlbaum Associates, Inc.

Novak, J. D. (1977). *A Theory of Education*. Ithaca, NY: Cornell University Press.

Osborne, R. J., & Whittrock, M. C. (1983). Learning science: A generative process. *Science Education*, *67*, 489-908.

Palmer, A.N. (2012). *Cave geology*. Laurel, MD: Cave Books.

Pan, Y. T., & Liu, S. C. (2018). Students' understanding of a groundwater system and attitudes towards groundwater use and conservation. *International Journal of Science Education*, *40*(5), 564-578.

Piaget, J. (1956). *The child's conception of space*. New York, NY: Macmillan Publishers.

Powell, K., & Kalina, C. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, *130*(2), 241–250.

Price, P. C., Jhangiani, R., & Chiang, I. C. A. (2015). *Research methods in psychology*. BCCampus.

Rees, C. (2018). Drawing on drawings: Moving beyond text in health professions education research. *Perspectives on medical education*, *7*(3), 166-173.

Reichard, J. (2010). *Environmental Geology*. Dubuque, IA: McGraw-Hill Higher Education.

Reinfried, S. (2006). Conceptual change in physical geography and environmental sciences through mental model building: The example of groundwater. *International Research in Geographical and Environmental Education*, *15*(1), 41-61.

Roberts, D. A., & Russell, T. L. (1975). An alternative approach to science education research: Drawing from philosophical analysis to examine practice. *Curriculum Theory Network*, *5*(2), 107-125.

Sapsford, R. (1999). *Survey research*. London: Sage Publishing.

Sinatra, G. M. & P. R. Pintrich. (2003). *Intentional conceptual change*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12*(4), 618-627.

Terada, Y. (2021, July 3). Why students forget – and what you can do about it. Edutopia. Retrieved from: https://www.edutopia.org/article/why-students-forget-and-what-you-can-do-about-it

United States Geological Survey. (2020, August 16). *Aquifer basics*. Retrieved from https://water.usgs.gov/ogw/aquiferbasics/

Unterbruner, U., Hilberg, S., & Schiffl, I. (2016). Understanding groundwater: Students' pre-conceptions and conceptual change by means of a theory-guided multimedia learning program. *Hydrology and Earth System Sciences*, *20*(6), 2251-2266.

Vosniadou, S. (2013). *International handbook of research on conceptual change*. New York, NY:Routledge.

Weary, D. J., & Doctor, D. H. (2014). *Karst in the United States: a digital map compilation and database*. US Department of the Interior, US Geological Survey.

Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science*, *14*(1), 6-9.

Wiggins, G., & McTighe, J. (2011). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Table 1

Table 1. Participant demographics compared against class and institutional demographics. (a) PKIS group drawn from two courses taught at one university. (b) Non-PKIS group drawn from a single course taught at another university. The column labeled Univ. undergrads lists the demographic data for the undergraduate population for each university at the time this study was conducted.

(a) Demographics for PKIS group		**Participants (n=61) %**	**Combined Course Enrollment (n=88) %**	**Univ. undergrads (n=20,081) %**
Gender	Male	44	49	53
	Female	56	51	47
Class Standing	Freshmen	37	34	19
	Sophomore	26	23	19
	Junior	20	20	25
	Senior	16	22	29
	Other	0	1	1
Race	Asian	10	3	2
	Caucasian	84	82	77
	Other	6	15	21
Major	STEM[1]	36	39	47
	Non-STEM	64	61	53
First generation[2]		13	16	9
International		8	9	8

(b) Demographics for Non-PKIS group		**Participants (n=55) %**	**Course Enrollment (n=59) %**	**Univ. undergrads (n=27,409) %**
Gender	Male	58	58	55
	Female	42	42	45
Class Standing	Freshmen	18	20	19
	Sophomore	38	37	25
	Junior	15	15	22
	Senior	29	27	25
	Other	0	0	8
Race	Asian	11	10	9
	Caucasian	71	73	67
	Other	18	17	24
Major	STEM[1]	44	42	44
	Non-STEM	56	58	56
First generation[2]		16	14	17
International		3	4	6

[1]STEM: science, technology, engineering, and mathematics

[2]Incoming first-generation students that academic year

Table 2

Table 2. Comparisons of conceptual understanding. (a) PKIS group and non-PKIS group at T_1. (b) PKIS group at T_1 and T_3. (c) Non-PKIS group at T_1 and T_3. (d) PKIS group and non-PKIS group at T_4. (e) PKIS group and non-PKIS group's learning gains from T_1 to T_3. (f) PKIS group at T_3 and T_4. (g) Non-PKIS group at T_3 and T_4.

(a)	PKIS T_1 Avg. Score (%)	PKIS T_1 Standard Error (%)	Non-PKIS T_1 Avg. Score (%)	Non-PKIS T_1 Standard Error (%)	t-value	p value	Cohen's d
	4.534	0.125	2.043	1.040	1.319	.190	0.259

(b)	PKIS T_1 Avg. Score (%)	PKIS T_1 Standard Error (%)	PKIS T_3 Avg. Score (%)	PKIS T_3 Standard Error (%)	t-value	p value	Cohen's d
	4.534	0.125	84.069	0.130	-26.573	< .00001*	5.2628

(c)	Non-PKIS T_1 Avg. Score (%)	Non-PKIS T_1 Standard Error (%)	Non-PKIS T_3 Avg. Score (%)	Non-PKIS T_3 Standard Error (%)	t-value	p value	Cohen's d
	2.043	1.040	65.505	4.374	-14.115	< .00001*	10.339

(d)	PKIS T_3 Avg. Score (%)	PKIS T_3 Standard Error (%)	Non-PKIS T_3 Avg. Score (%)	Non-PKIS T_3 Standard Error (%)	t-value	p value	Cohen's d
	84.069	0.130	65.505	4.374	3.653	< .001*	0.722

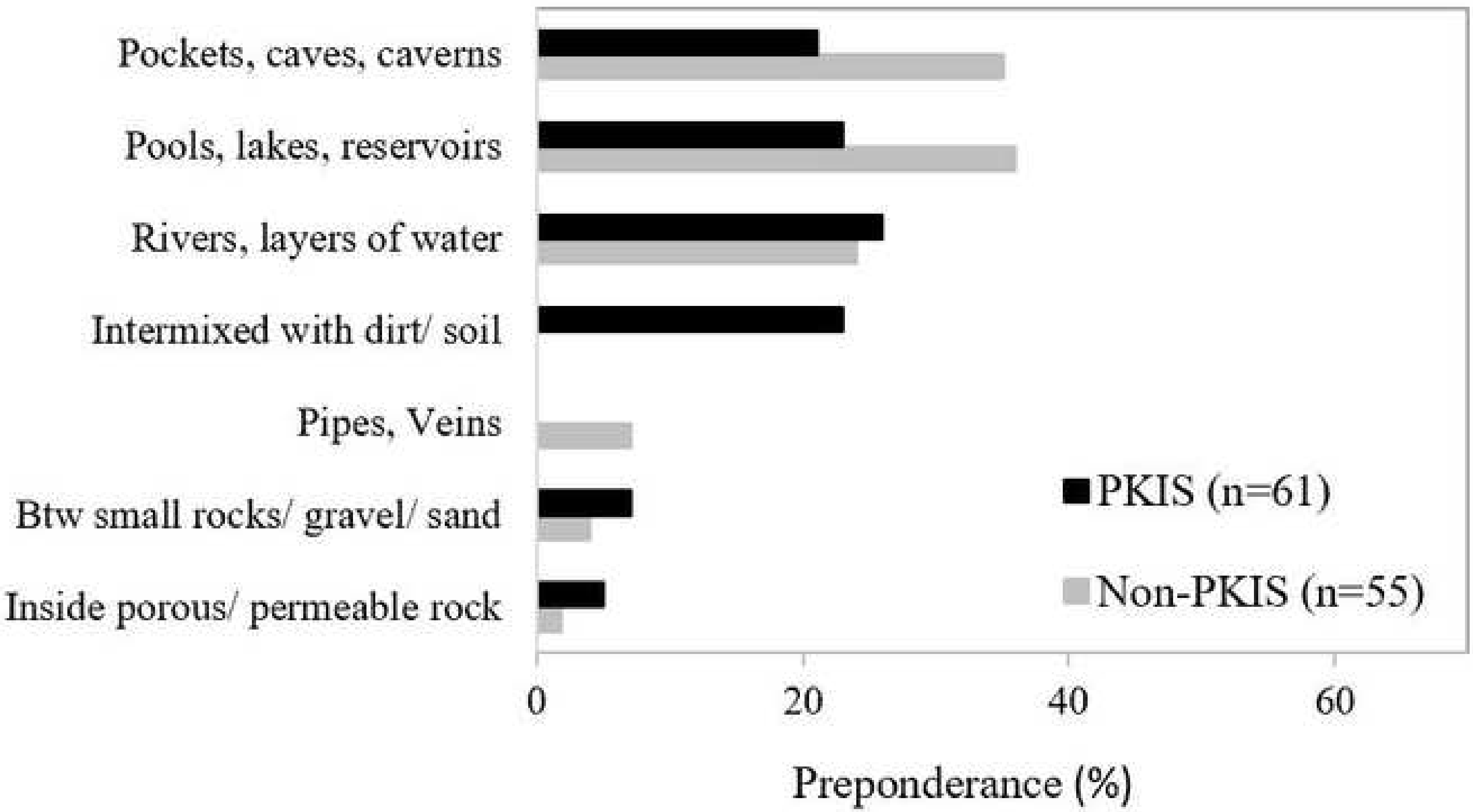
(e)	PKIS <g> Avg. Score	PKIS <g> Standard Error	Non-PKIS <g> Avg. Score	Non-PKIS <g> Standard Error	t-value	p value	Cohen's d
	0.817	0.035	0.643	0.045	2.898	.005*	0.608

(f)	PKIS T_3 Avg. Score (%)	PKIS T_3 Standard Error (%)	PKIS T_4 Avg. Score (%)	PKIS T_4 Standard Error (%)	t-value	p value	Cohen's d
	84.069	0.130	83.088	3.343	0.220	.826	0.0436

(g)	Non-PKIS T_3 Avg. Score (%)	Non-PKIS T_3 Standard Error (%)	Non-PKIS T_4 Avg. Score (%)	Non-PKIS T_4 Standard Error (%)	t-value	p value	Cohen's d
	65.505	4.374	49.760	5.195	2.319	.022*	0.455

* $p < .05$

Figure 1 TIF



Pockets, caves, caverns
Pools, lakes, reservoirs
Rivers, layers of water
Intermixed with dirt/ soil
Pipes, Veins
Btw small rocks/ gravel/ sand
Inside porous/ permeable rock
PKIS (n=61)
Non-PKIS (n=55)
0
20
40
60
Preponderance (%)

Figure 1 with caption

Figure 1. Preconceptions of groundwater residence held by the PKIS group and non-PKIS group. The sum for each group is greater than 100% because more than one mental model could be depicted in the same sketch.

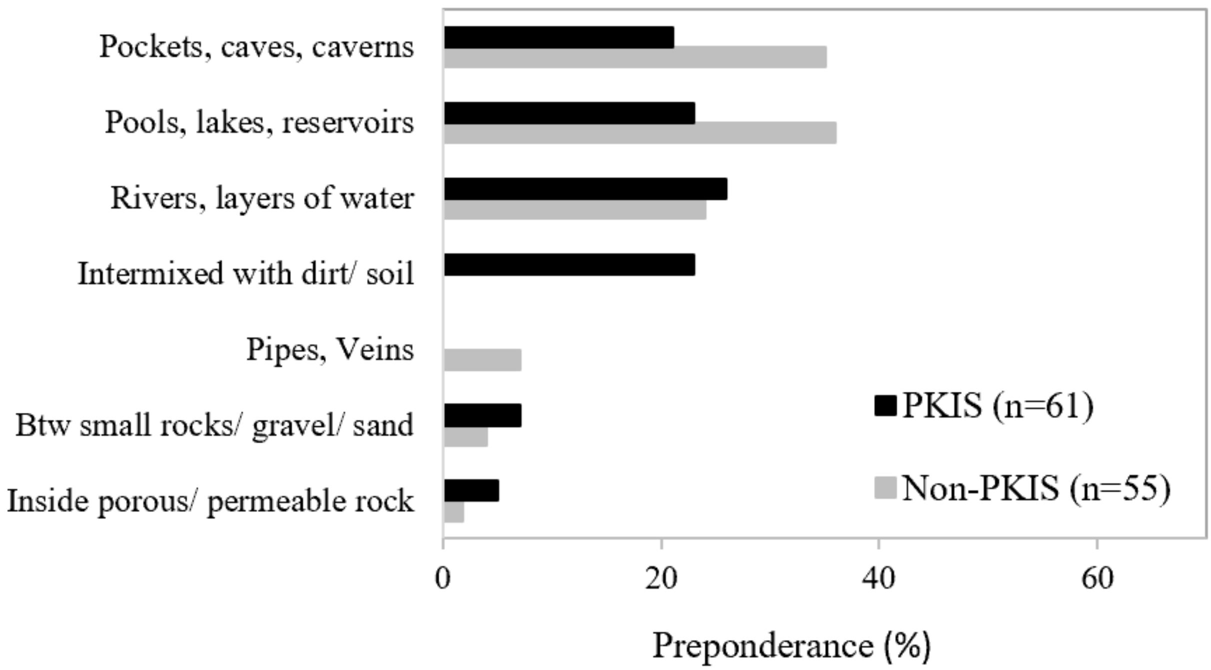
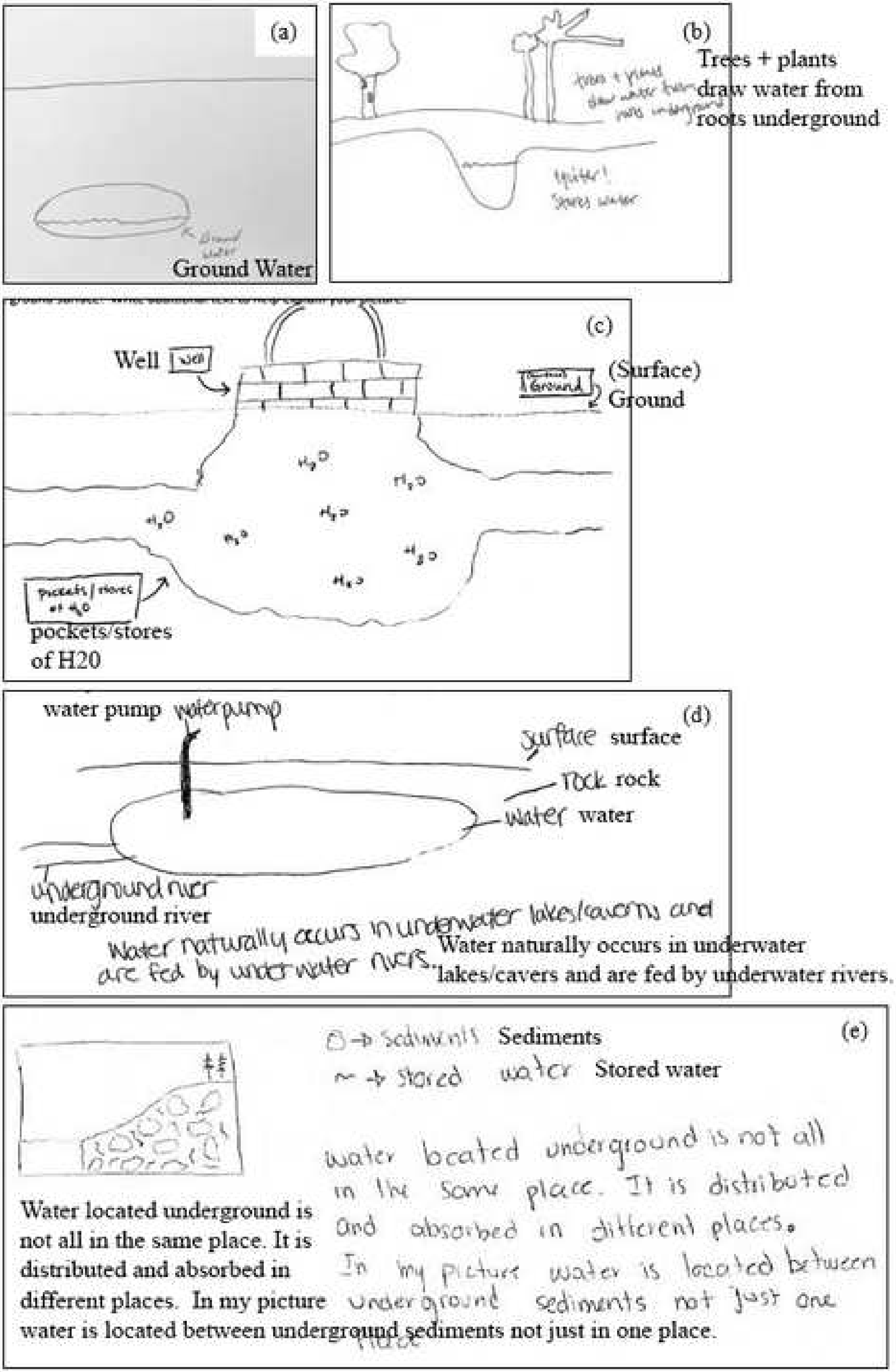


Figure 2 TIF



(a)
Ground Water
(b)
Trees + plants draw water from roots underground
(c)
Well
(Surface) Ground
pockets/stores of H20
(d)
water pump
surface
rock
water
underground river
Water naturally occurs in underwater lakes/cavers and are fed by underwater rivers.
(e)
Sediments
Stored water
Water located underground is not all in the same place. It is distributed and absorbed in different places. In my picture water is located between underground sediments not just in one place.

Figure 2 with caption

Figure 2. Annotated sketches of pre-instruction conceptions of groundwater residence. (a)-(d) Examples of mental model for continuous body of water. (e) Example of water stored between sediments.

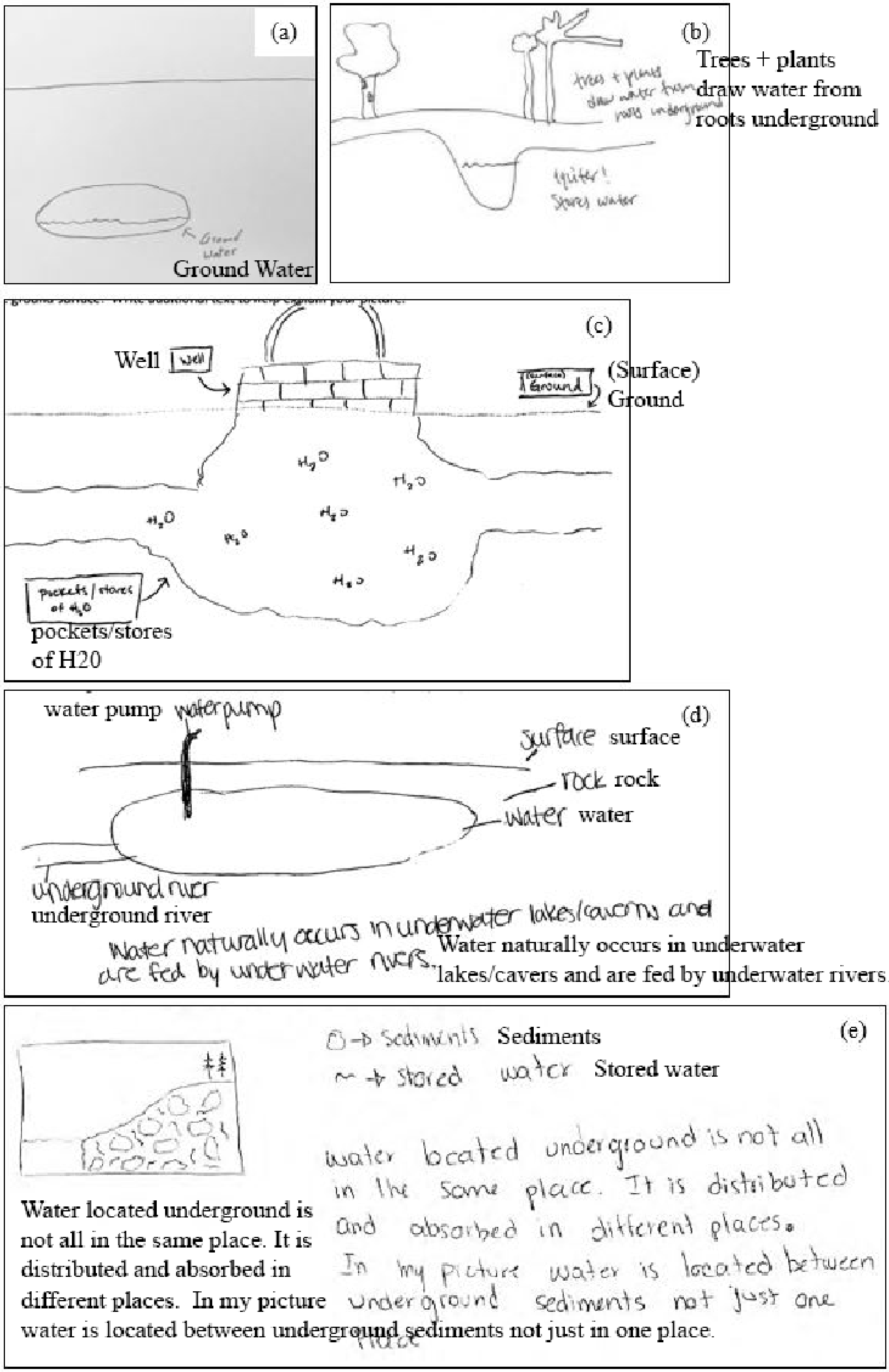
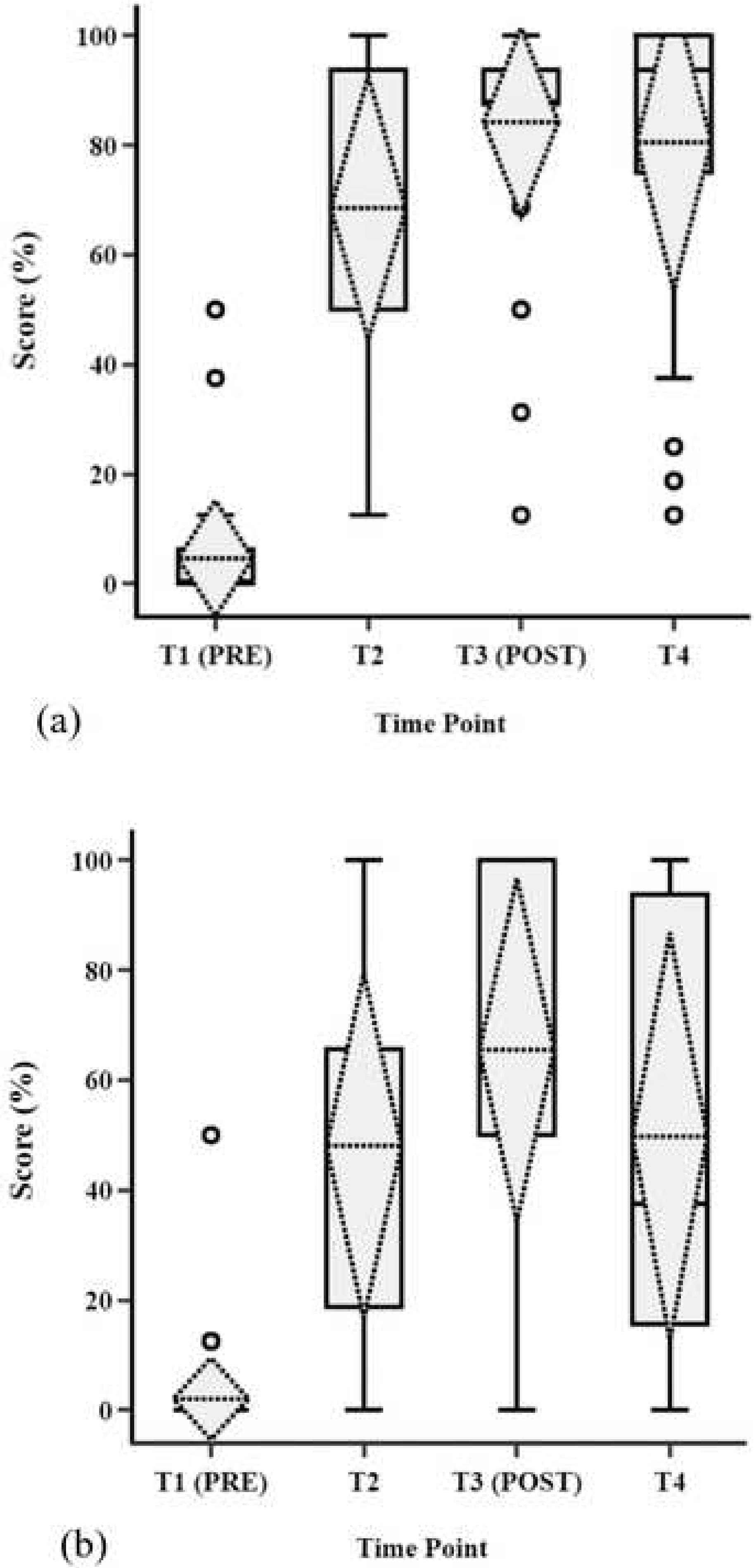


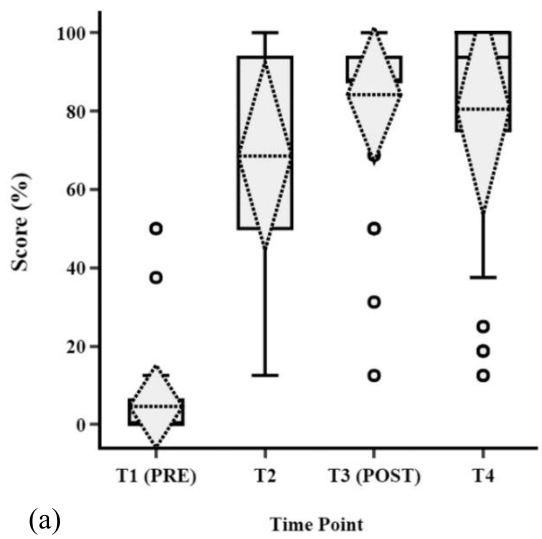
Figure 3



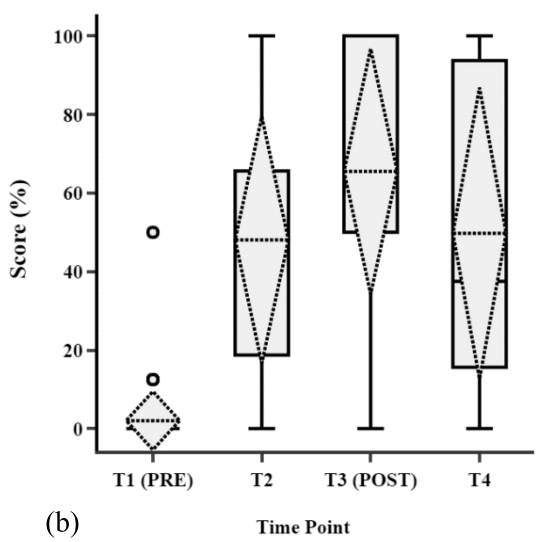
Score (%)
100
80
60
40
20
0
T1 (PRE)
T2
T3 (POST)
T4
(a)
Time Point
Score (%)
100
80
60
40
20
0
T1 (PRE)
T2
T3 (POST)
T4
(b)
Time Point

Figure 3 with caption

Figure 3. Conceptual change over time. (a) PKIS group. (b) Non-PKIS group. The dashed line represents the mean, and the dashed triangles represent the standard deviation. The circles indicate outliers. As is typical with box-and-whisker plots, the upper-bound of the rectangle represents the upper quartile, the lower-bound of the rectangle represents the lower quartile, the solid vertical line between them represents the median, the upper horizontal line on the whisker represents the max observed value or upper fence, and the lower horizontal line on the whisker represents the min observed value or lower fence. This visualization was created using BioVinci version 1.1.5 developed by BioTurning Inc., San Diego California USA, www.bioturing.com



(a)



(b)

Figure 4

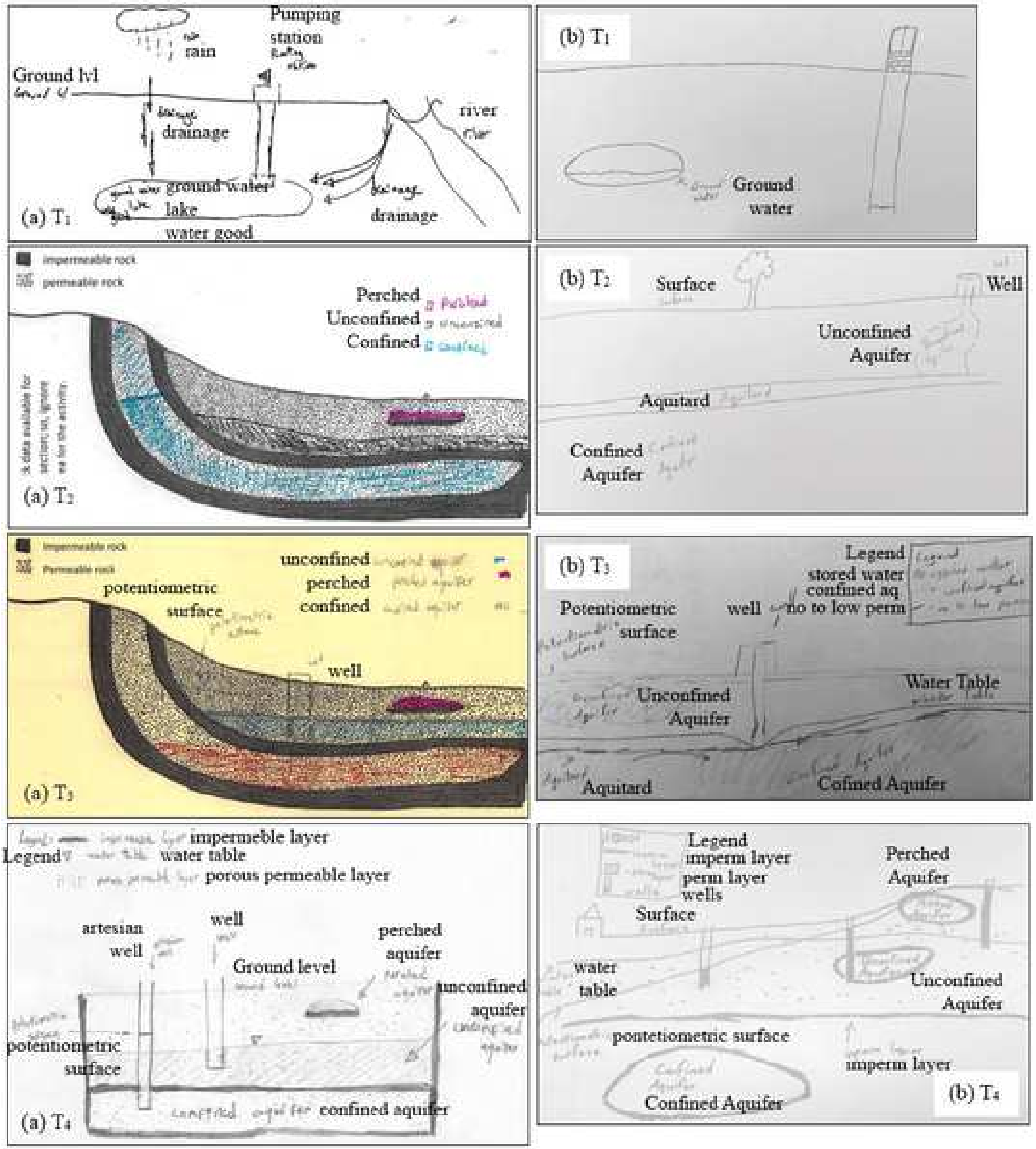

Pumping station
rain
Ground lvl
drainage
river
ground water
lake
water good
drainage
(a) T1
(b) T1
Ground water
impermeable rock
permeable rock
Perched
Unconfined
Confined
(a) T2
(b) T2
Surface
Well
Unconfined Aquifer
Aquitard
Confined Aquifer
impermeable rock
Permeable rock
unconfined
perched
confined
potentiometric surface
well
(a) T3
(b) T3
Legend
stored water
confined aq.
no to low perm
Potentiometric surface
well
Unconfined Aquifer
Water Table
Aquitard
Cofined Aquifer
Legend
impermeble layer
water table
porous permeable layer
artesian well
well
Ground level
perched aquifer
unconfined aquifer
potentiometric surface
confined aquifer
(a) T4
Legend
imperm layer
perm layer
wells
Surface
Perched Aquifer
water table
Unconfined Aquifer
pontetiometric surface
imperm layer
Confined Aquifer
(b) T4

Figure 4 with caption

Figure 4. Sketches illustrating conceptual change over time for (a) one student in the PKIS group and (b) one student in the non-PKIS group.

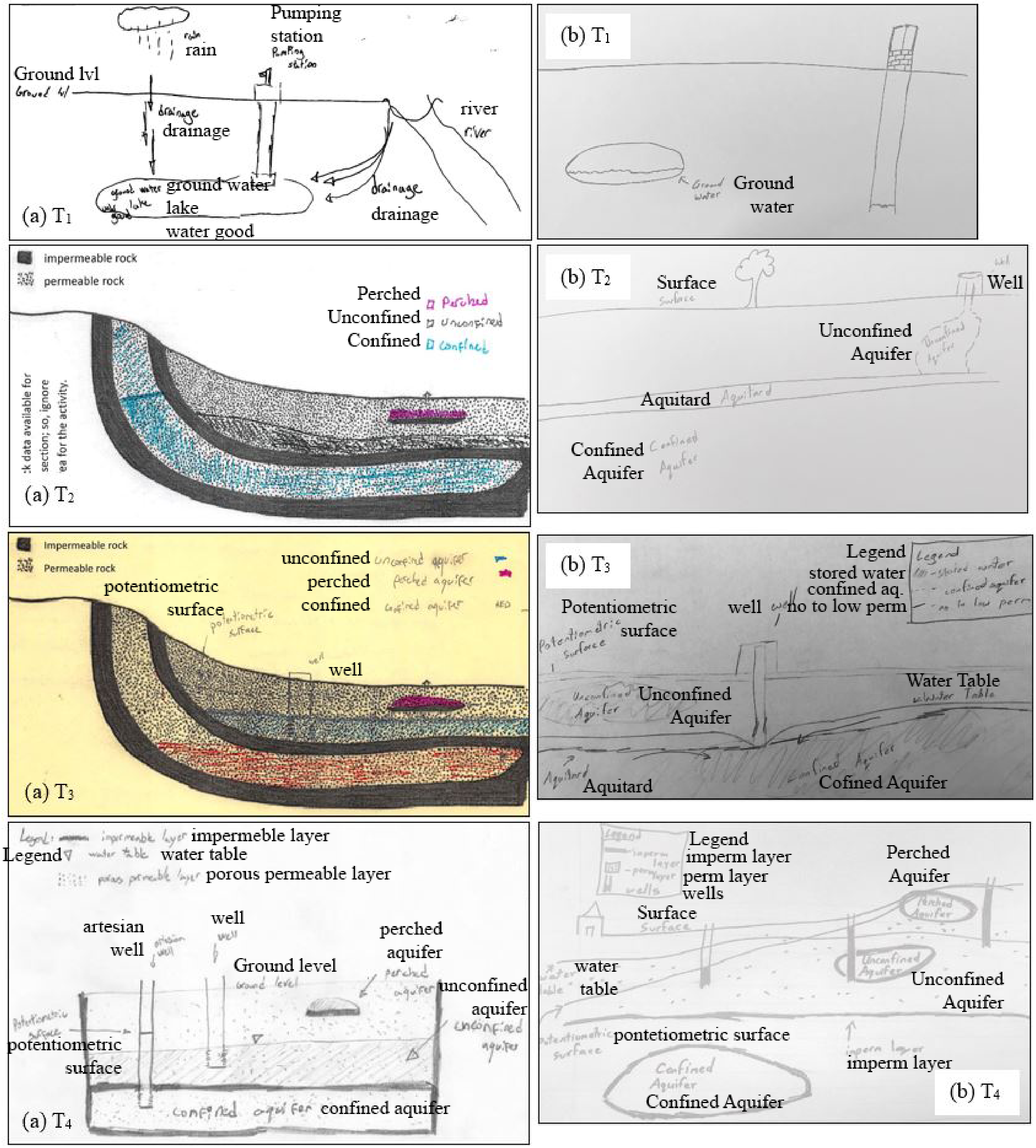


Table S1. Sampling timeline relative to week in the semester and course content. Both courses are introductory-level geoscience courses that satisfy the natural science requirement for graduation. Both are survey courses that address a different topic each week. T_2 sampling occurs in both courses during the week that students learned about aquifers, groundwater residence, and groundwater flow. T_1 sampling in the PKIS group and the non-PKIS group occurred before groundwater-related concepts were first addressed in both courses. T_2 sampling occurred during the week where both curses had instruction on aquifers and groundwater flow. T_3 sampling occurred two to three weeks after the instructional sequence of interest. The PKIS group received one week of direct instruction related to groundwater concepts, whereas the non-PKIS group had about six weeks of direct instruction on groundwater concepts. Time points T_1, T_2, and T_3 sampling data are used to make inter-group comparisons. Time point T_4 sampling data are used only to make within-group comparisons because of the large difference in time that elapsed between T_3 and T_4 in the non-PKIS group compared to the PKIS group.

Week in Semester	PKIS time point	PKIS Group course topic	Non-PKIS time point	Non-PKIS Group course topic
1		What is science and how do scientists know what they know	T_1	Rocks and minerals
2		Volcanoes		Fluid storage and mobility
3		Earthquakes		Water cycle and human water uses
4		Plate tectonics	T_2	Aquifers and groundwater resources
5		Waves and floods		Impacts and extending water supply
6		Hurricanes and tornadoes	T_3	Water quality
7		Rock and mineral resources		Water remediation
8	T_1	Soil and forest resources		Energy sources
9	T_2	Aquifers and groundwater resources		Petroleum systems
10		Fossil fuel resources		Unconventional hydrocarbons
Semester break				
11	T_3	Non/Renewable resources		Geothermal energy
12		Mining		Solar and nuclear energy
13		Pollution and waste		Mining
14		Climate change		Impacts of mining
15		Possible solutions		Mining remediation
16	T_4		T_4	

Table S2. Prompts used to ask students sketch their ideas about groundwater and how it is naturally stored underground.

Group	Time Point	Prompt	Type of Item
Non-PKIS	T_1	Draw and label a picture of how water (used for drinking water) naturally exists or is naturally stored below the ground surface. Write additional text to help explain your answer.	Free-sketch in-class activity
Non-PKIS	T_2	Draw and label a picture of how water (used for drinking water) naturally exists or is naturally stored below the ground surface. Write additional text to help explain your answer.	Free-sketch in-class activity
Non-PKIS	T_3	Draw and label a cross-section of the subsurface that shows the position of the following geologic structures or features: confined aquifer, unconfined aquifer, aquitard, water table, potentiometric surface, direction of groundwater flow, and one or more wells as needed to illustrate certain structures and/or relationships in your cross-section.	Free-sketch mid-term exam item
Non-PKIS	T_4	Draw and label a cross-section that shows the position of the following geologic structures with respect to one another: confined aquifer, unconfined aquifer, perched aquifer; water table, potentiometric surface; impermeable layers, porous and permeable layers; and wells as needed to illustrate certain hydrological relationships.	Free-sketch final exam item
PKIS	T_1	In preparation for next week, draw and label a picture of how water * is naturally stored below the ground. *water that is pumped from the ground to drink	Free-sketch in-class activity
PKIS	T_2	How are all three aquifers related to each other in a 'bigger picture'? Let's give it a try! Use color pencils if you brought some. On your handout: Shade in where each of the three types of aquifers would occur. Be sure to label each aquifer that you shaded in. Note: There is a little house sketched in for reference, to help you visualize the size and extent of the aquifers.	Delimited-sketch in-class activity
PKIS	T_3	In the figure below, (1) draw in the confined, perched, and unconfined aquifers; (2) draw in a drinking water well that pumps water out of the unconfined aquifer; (3) label each aquifer, the water table, ad the potentiometric surface; ad (4) in the space below, answer the following question: 'What does it mean for a rock to be impermeable'? by completing the sentence: For a rock to be considered impermeable, it means that …	Delimited-sketch mid-term exam item
PKIS	T_4	Draw and label a sketch that shows the position of the following geologic features with respect to one another: confined aquifer, unconfined aquifer, perched aquifer, water table, potentiometric surface, impermeable layers, porous and permeable layers, and wells as needed to illustrate certain relationships.	Free-sketch final exam item

Table S3. Prompts posed in the pre-course and post-course surveys to gain insights about the social dimension of learning.

Prompt	Type of Item
I like to work alone. Likert-scale answer choices: (1) Strongly disagree, (2) Disagree, (3) Unsure, (4) Agree, (5) Strongly agree	Likert-scale
I like to work with others. Likert-scale answer choices: (1) Strongly disagree, (2) Disagree, (3) Unsure, (4) Agree, (5) Strongly agree	Likert-scale
I strongly prefer to work … Multiple-choice answer choices: (1) alone, (2) with others, (3), both, (4) no preference	Multiple-choice
What did you like best about this course?	Open-ended
What can the instructor do to improve the course when she teaches it in the future?	Open-ended

Table S4. Scoring rubric applied to students' concept sketches.

Aquifer Type	Location Criteria	0 Point	0.25 Point	0.5 Point	1 Point	1.5 Points	1.75 Points	2 Points
Perched	Water-saturated permeable rock layer that rests on an impermeable rock lens above the water table or an unconfined aquifer	Not shown in T_1 and T_2. Not shown or not labeled in T_3 and T_4.	Incorrect location (e.g., inside unconfined aquifer) but labeled something. Water body encased in impermeable rock shell.	Partially correct. Above unconfined aquifer but does not rest on an impermeable rock lens.	Correct location but incorrect lateral extent (e.g., drawn as a 'packet' or points at general location0.	NA	NA	Correctly positioned in terms of location and lateral extent.
Unconfined	Water-saturated permeable rock layer that rests on an impermeable rock layer and has vertical pore connection to the atmosphere	Not shown in T_1 and T_2. Not shown or not labeled in T_3 and T_4.	Incorrect location (e.g., envelopes a confined aquifer) but labeled something. Water body encased in impermeable rock shell.	Partially correct. Below the unsaturated zone but does not rest on an impermeable rock layer.	Correct location but incorrect lateral or vertical extent (e.g., drawn as a 'packet' or points at general location). The water table is not the top boundary of the unconfined aquifer. The water table has thickness and is drawn as a layer.	NA	Mostly correct location but the vertical extent abuts impermeable rock, thus creating a confined aquifer.	Correctly positioned in terms of location as well as lateral and vertical extent.
Confined	Water-saturated permeable rock layer that rests between two layers of impermeable rock and has no vertical pore connection to the atmosphere	Not shown in T_1 and T_2. Not shown or not labeled in T_3 and T_4.	Incorrect location but labeled something. Water body encased in impermeable rock shell.	Partially correct. Has a top impermeable rock layer but no bottom impermeable rock layer (or vice versa).	Correct location but incorrect lateral extent (e.g., drawn as a 'packet' or points between two impermeable rock layers) and/or incorrect vertical extent (e.g., groundwater does not fill vertical space in the permeable rock between the two impermeable rock layers).	Correct or mostly correct lateral extent, below unconfined aquifer, but missing bottom impermeable rock layer.	Correct location and mostly correct lateral extent.	Correctly positioned in terms of location as well as lateral and vertical extent.

Table S5. Homework, mid-term exam, and final exam questions in both courses provide supplemental insight into students' learning progress; however, they are not directly comparable between the PKIS and non-PKIS groups as the two courses asked different homework, mid-term exam, and final exam questions. Questions most directly related to groundwater residence and aquifers in the course from which the PKIS group were enrolled are included below.

Assessment Type*	Item	PKIS, % Correct (n=61)
HW	Karst terrain is not widespread in the US. However, Florida is a US state that is characterized by karst terrain in which subterranean caves can form. Explain under what conditions such a subterranean cave could be considered an aquifer. (free-response)	76
HW	An aquifer composed of ____ will have the greatest porosity compared to aquifers of the other materials listed. (multiple-choice)	82
HW	Apply your knowledge of rock types, porosity, permeability, and aquifers to answer the following question. Which one of the following materials could hold a lot of water but would not allow water to flow through it readily (assuming all other factors are held equal)? (multiple-choice)	56
HW	____ has very low permeability, whereas ____ has very high permeability. (multiple-choice)	89
HW	Examine the relationship between the unconfined and perched aquifers shown in the figure provided. (a) Say what is incorrect with the perched aquifer. (b) Pull from what was discussed in class and the textbook to provide the rationale/reason for why it is incorrect. (free-response)	69
ME	Which one of the following is true about the saturated zone? (multiple-choice)	91
ME	Which of the aquifers listed below is charged locally? (multiple-choice)	85
FE	Generally, when over pumping of groundwater occurs in inland regions far from the coast, _____ around the well. (multiple-choice)	95
FE	Generally, when over pumping of groundwater occurs in coastal communities, _____ around the well. (multiple-choice)	94

* HW = homework at end of instructional sequence, ME = mid-term exam three weeks after end of instructional sequence, FE = final exam eight weeks after end of instructional sequence

Table S6. Homework, mid-term exam, and final exam questions in both courses provide supplemental insight into students' learning progress; however, they are not directly comparable between the PKIS and non-PKIS groups as the two courses asked different homework, mid-term exam, and final exam questions. Questions most directly related to groundwater residence and aquifers in the course from which the non-PKIS group were enrolled are included below.

Assessment Type*	**Item**	**Non-PKIS, % Correct (n=55)**
HW	What are the defining characteristics of an *aquifer*? (free-response)	61
HW	What are the defining characteristics of an *aquitard*? (free-response)	87
HW	What is the difference between an *unconfined aquifer* and a *confined aquifer*? (free-response)	77
HW	Use Figure 11.2 on page 24 in your textbook (or the figure provided) to (a) name one sediment or one sedimentary rock type that makes a good *aquifer* and (b) report its approximate permeability (in units of m^2). (free-response)	87
HW	Use Figure 11.2 on page 24 in your textbook (or the figure provided) to (a) name one sediment or one sedimentary rock type that makes a good *aquitard* and (b) report its approximate permeability (in units of m^2). (free-response)	79
ME	What is permeability? (multiple-choice)	92
ME	Below are thin sections for two sandstones. In both thin sections, pores are orange in color and clasts are white in color (and a few clasts are blue in Sandstone A). Which sandstone has greater permeability? (multiple-choice)	90
ME	The figure below shows that loose sediments exhibit a wide range of porosity and permeability. Why do loose sediments exhibit a wide range of porosity and permeability? (multiple-choice)	92
FE	The final exam did not include questions about groundwater residence and aquifers, except for the T_4 sketch question posed to compare the PKIS and non-PKIS for the purposes of this study.	NA

* HW = homework at end of instructional sequence, ME = mid-term exam three weeks after end of instructional sequence, FE = final exam eight weeks after end of instructional sequence. NA = not applicable

Table S7. Temporal data comparing the PKIS Group's performance by (a) gender and (b) race. There are neither statistically significant differences between the performance of male and female students nor between Caucasian and non-Caucasian students.

(a) PKIS Group gender	T1. Pre-Instruction sample (n)	T1. Pre-Instruction Score (median %)	T1. Pre-Instruction Score (average %)	T1. Pre-Instruction Score (stdev)	T1. Pre-Instruction Sampling Error (%)	T2. Instruction sample (n)	T2. Instruction Score (median%)	T2. Instruction Score (average%)	T2. Instruction Score (stdev)	T2. Instruction Sampling Error (%)	T3. Instruction sample (n)	T3. Instruction Score (median%)	T3. Instruction Score (average%)	T3. Instruction Score (stdev)	T3. Instruction Sampling Error (%)	T4. Post- Instruction sample (n)	T4. Post-Instruction Score (median%)	T4. Post-Instruction Score (average%)	T4. Post-Instruction Score (stdev)	T4. Post-Instruction Sampling Error (%)	T4-T1 change in average (%)
female	34	0.00	2.94	8.18	1.40	40	37.50	52.59	28.60	4.52	61	95.83	82.68	22.58	2.89	34	95.83	79.66	28.66	4.92	76.72
male	27	0.00	4.32	7.79	1.50	25	50.00	59.00	31.61	6.32	25	95.83	84.83	20.12	4.02	27	95.83	86.88	19.77	3.80	82.56
overall	61	0.00	3.55	7.97	1.02	61	37.50	55.56	29.92	3.83	61	95.83	83.63	21.37	2.74	61	95.83	82.86	25.19	3.22	79.30
t-value			-0.668					-0.783					-0.374					-1.115			
p value			0.506					0.437					0.710					0.269			
significant?			no					no					no					no			

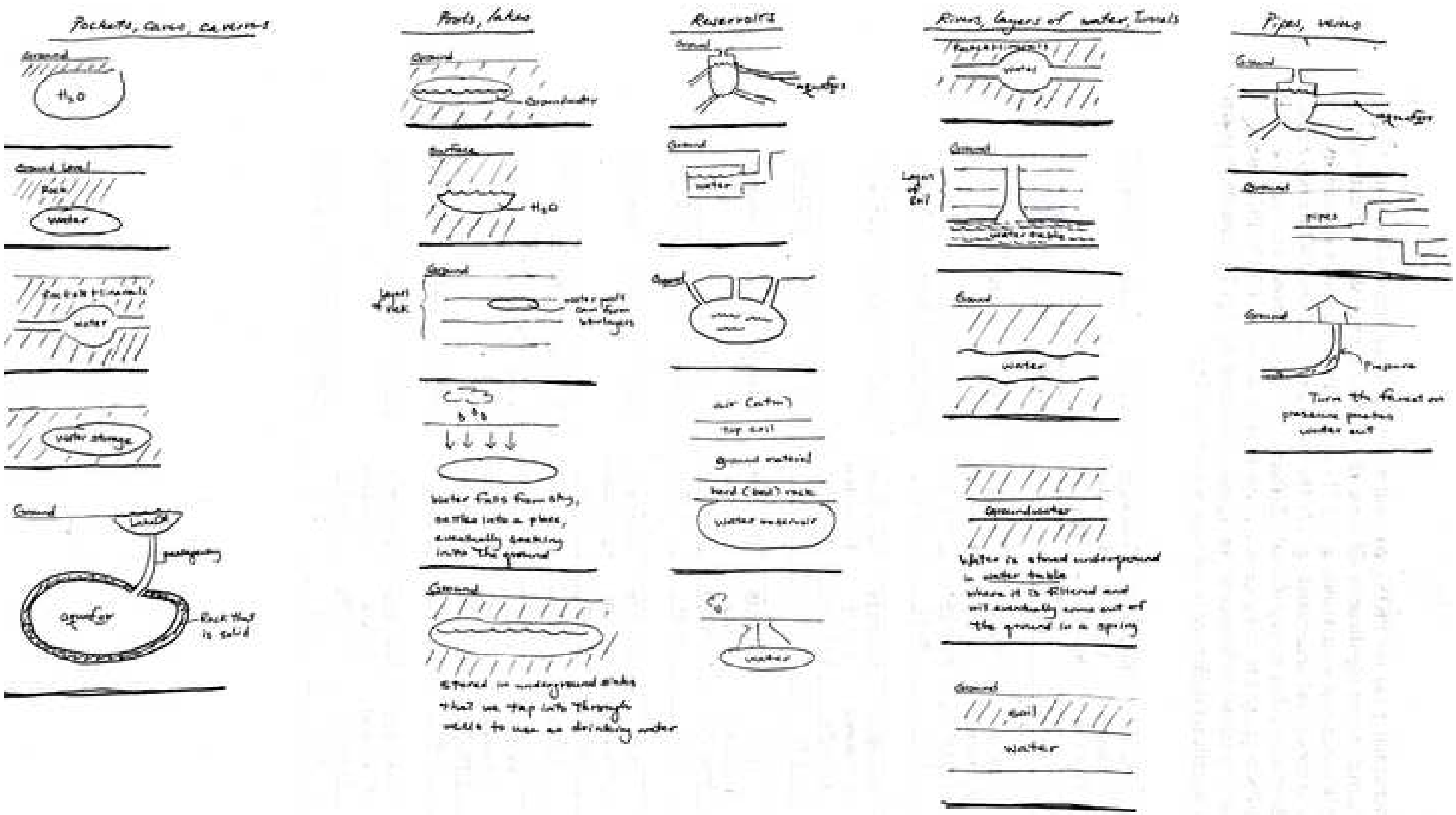
(b) PKIS Group race	T1. Pre-Instruction sample (n)	T1. Pre-Instruction Score (median %)	T1. Pre-Instruction Score (average %)	T1. Pre-Instruction Score (stdev)	T1. Pre-Instruction Sampling Error (%)	T2. Instruction sample (n)	T2. Instruction Score (median%)	T2. Instruction Score (average%)	T2. Instruction Score (stdev)	T2. Instruction Sampling Error (%)	T3. Instruction sample (n)	T3. Instruction Score (median%)	T3. Instruction Score (average%)	T3. Instruction Score (stdev)	T3. Instruction Sampling Error (%)	T4. Post- Instruction sample (n)	T4. Post-Instruction Score (median%)	T4. Post-Instruction Score (average%)	T4. Post-Instruction Score (stdev)	T4. Post-Instruction Sampling Error (%)	T4-T1 change in average (%)
Caucasian	51	0.00	2.78	6.16	0.86	40	37.50	57.10	29.25	4.62	61	95.83	83.60	21.27	2.72	34	95.83	80.47	26.87	4.61	77.70
non-Caucsn	10	0.00	7.50	13.86	4.38	25	35.42	48.75	33.45	6.69	25	93.75	83.75	23.03	4.61	27	95.83	95.00	4.73	0.91	87.50
overall	61	0.00	3.55	7.97	1.02	61	37.50	55.56	29.92	3.83	61	95.83	83.63	21.37	2.74	61	95.83	82.86	25.19	3.22	79.30
t-value			-1.741					0.794					-0.020					-1.693			
p-value			0.087					0.215					0.984					0.096			
significant?			no					no					no					no			

Table S8. Temporal data comparing the non-PKIS Group's performance by (a) gender and (b) race. There were neither statistically significant differences between the performance of male and female students nor between Caucasian and non-Caucasian students. Only students who completed both the T_1 and T_4 assessments were included in this analysis. A few of them did not include the T_2 and T_3 assessments, which explains why the totals for study participants during T_2 and T_3 might not equal those for T_1 and T_4.

(a) Non-PKIS Group gender	T1. Pre-Instruction sample (n)	T1. Pre-Instruction Score (median %)	T1. Pre-Instruction Score (average %)	T1. Pre-Instruction Score (stdev)	T1. Pre-Instruction Sampling Error (%)	T2. Instruction sample (n)	T2. Instruction Score (median%)	T2. Instruction Score (average%)	T2. Instruction Score (stdev)	T2. Instruction Sampling Error (%)	T3. Instruction sample (n)	T3. Instruction Score (median%)	T3. Instruction Score (average%)	T3. Instruction Score (stdev)	T3. Instruction Sampling Error (%)	T4. Post- Instruction sample (n)	T4. Post-Instruction Score (median%)	T4. Post-Instruction Score (average%)	T4. Post-Instruction Score (stdev)	T4. Post-Instruction Sampling Error (%)	T4-T1 change in average (%)
female	23	0.00	3.26	10.80	2.25	21	50.00	39.58	31.02	6.77	23	62.50	64.95	33.11	6.90	23	68.75	58.97	35.80	7.46	55.71
male	32	0.00	1.37	3.45	0.61	32	50.00	52.54	31.66	5.60	31	62.50	62.30	32.97	5.92	32	25.00	44.14	38.30	6.77	42.77
overall	55	0.00	2.16	10.38	1.40	53	50.00	47.41	31.76	4.36	55	62.50	63.43	32.75	4.46	55	37.50	50.34	37.66	5.08	48.18
t-value			0.930					-1.469					0.291					1.455			1.217
p value			0.356					0.148					0.772					0.152			0.229
signifcant?			no					no					no					no			no

(b) Non-PKIS Group race	T1. Pre-Instruction sample (n)	T1. Pre-Instruction Score (median %)	T1. Pre-Instruction Score (average %)	T1. Pre-Instruction Score (stdev)	T1. Pre-Instruction Sampling Error (%)	T2. Instruction sample (n)	T2. Instruction Score (median%)	T2. Instruction Score (average%)	T2. Instruction Score (stdev)	T2. Instruction Sampling Error (%)	T3. Instruction sample (n)	T3. Instruction Score (median%)	T3. Instruction Score (average%)	T3. Instruction Score (stdev)	T3. Instruction Sampling Error (%)	T4. Post- Instruction sample (n)	T4. Post-Instruction Score (median%)	T4. Post-Instruction Score (average%)	T4. Post-Instruction Score (stdev)	T4. Post-Instruction Sampling Error (%)	T4-T1 change in average (%)
Caucasian	16	0.00	4.69	12.60	3.15	14	50.00	45.00	31.27	8.36	16	62.50	55.47	40.56	10.14	16	28.13	46.47	40.61	10.15	41.78
non-Caucsn	39	0.00	1.12	3.47	0.56	39	50.00	48.36	32.32	5.18	38	62.50	66.78	28.81	4.67	39	56.25	53.13	36.77	5.89	52.00
overall	55	0.00	2.16	7.44	1.00	53	50.00	47.41	31.76	4.36	54	62.50	63.43	32.75	4.46	55	37.50	50.34	37.66	5.08	48.18
t-value			1.640					-0.344					-1.163					-0.581			-0.869
p-value			0.107					0.733					0.250					0.563			0.389
significant?			no					no					no					no			no

SM Figure S1 - TIF



SM Figure S1 with caption

Figure S1. Rubric for pre-instructional mental models, based on categories that emerged from analyzing students' sketches. The emergent categories include 'pockets, caves, and caverns'; 'pools and lakes'; 'reservoirs'; 'rivers and layers of water', and 'pipes and veins.' Each category includes author-recreated student sketches to illustrate the range in sketches for each emergent category.