



Designing Grounded Feedback: Criteria for Using Linked Representations to Support Learning of Abstract Symbols

Eliane S. Wiese¹ · Kenneth R. Koedinger²

Published online: 13 January 2017

© International Artificial Intelligence in Education Society 2016

Abstract This paper proposes *grounded feedback* as a way to provide implicit verification when students are working with a novel representation. In grounded feedback, students' responses are in the target, to-be-learned representation, and those responses are reflected in a more-accessible linked representation that is intrinsic to the domain. By examining the accessible feedback representation, students can infer if their work with the novel representation is correct. This paper presents the criteria for grounded feedback, provides examples of systems that implement grounded feedback, contrasts grounded feedback with similar feedback types, and discusses the evidence for grounded feedback's effectiveness. Controlled experiments with random assignment that compare grounded feedback to other approaches are limited in number and scope (i.e., comparisons to explicit verification with and without text hints, linked representations, and no feedback). The two experiments we found with full implementation of grounded feedback and a sample size larger than 20 found robust learning benefits of grounded feedback over explicit verification feedback. These results are promising and indicate that grounded feedback warrants further investigation.

Keywords Grounded feedback · Feedback · Multiple representations · Designing interactive learning environments

✉ Eliane S. Wiese
eliane.wiese@berkeley.edu

Kenneth R. Koedinger
koedinger@cmu.edu

¹ Graduate School of Education, University of California, Berkeley, CA, USA

² Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Introduction: Feedback for Understanding Symbols and the Procedures that Use Them

Science, technology, engineering, and math (STEM) domains include systems of abstract symbols and procedures that use them. How can feedback best support learning how to use these symbols and perform procedures with them, while also recognizing the conceptual principles that underlie them, and the appropriate application of those concepts (Schoenfeld 1988)? Students may learn the concepts underlying these abstract symbols more easily when the unfamiliar symbols are connected to already-familiar representations. This may make relevant features more salient and, thus, easier to reason with. For example, a student may think that $1/10$ is greater than $1/4$, since ten is greater than four. When the two fractions are plotted on a number line, their magnitudes become more salient, and the student should recognize that $1/4$ is greater.

This paper presents *grounded feedback* as a way to leverage an accessible representation to help students make sense of a novel one. This paper synthesizes past work to propose candidate guidelines for design. These guidelines are theoretically motivated and have some existing empirical evidence for their effectiveness. This paper identifies common feedback features in previous work, synthesizes them into four criteria, provides examples of how to implement grounded feedback in existing systems, and presents current evidence on the effectiveness of grounded feedback.

Many intelligent tutoring systems make use of multiple representations to support students in learning abstract concepts, from physics and chemistry (Dzikovska et al. 2014; Rau and Evenstone 2014) to arithmetic (Käser et al. 2013; Pareto 2014). In some cases, all of the graphical representations are important to learn in themselves (Rau and Evenstone 2014), while in other cases a concrete representation is used as a stepping-stone (Käser et al. 2013; Pareto 2014). Technology offers many possibilities for learning with multiple representations: representations can be static, dynamic, interactive, and shown as feedback (among other options). When the ultimate goal is fluency with a less-accessible representation, how should instructional designs leverage a more-accessible representation? One promising approach is to use representations that support qualitative thinking, such as strip diagrams in Singapore Math (Fig. 1; Beckmann 2004). Such diagrams are not intended to help students execute symbolic procedures, but rather are intended to support students in qualitative reasoning (e.g., Which amounts are bigger?) and planning (e.g., Which operation is needed?). *Grounded feedback* extends this idea of using an additional representation to support reasoning. To highlight the features of grounded feedback, we contrast it with similar approaches: explicit verification feedback

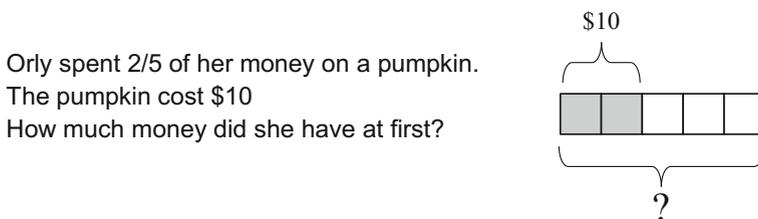


Fig. 1 Strip diagrams support qualitative reasoning (Orly had more than \$10 at first) and planning (find half of \$10, then multiply that amount by 5 to find the original amount). However, as static representations, strip diagrams do not provide feedback on students' algebraic symbolization

(indicating if a response is correct or incorrect; Shute 2008), static representations, and linked representations. While experiments comparing grounded feedback to explicit verification feedback have shown benefits for grounded feedback, we have not found experiments with sample sizes above 20 comparing grounded feedback to other uses of multiple representations.

Definition of Grounded Feedback

Grounded feedback shows the student two representations: the target symbolic representation and a feedback representation that is more accessible (often more concrete or more familiar). Students solve problems with the symbolic representation, and their answers are visualized with the feedback representation. Grounded feedback must be matched to the prior knowledge of the students such that they can anticipate what a correct answer would look like in the feedback representation but not in the target symbolic representation. Therefore, grounded feedback allows students to evaluate if their answers are correct, without showing explicit verification or the correct response with the symbolic representation. Instead, students use the more accessible representation to infer which aspects of their symbolic answer are correct or incorrect.

Criteria of Grounded Feedback

An important aim of grounded feedback is to help students decide if their work is correct. Therefore, the first criterion of grounded feedback is that *with the feedback, students can easily identify a correct answer*. Additionally, for this criterion to be met, identifying a correct answer must be easier with the feedback than without it. A simple way to fulfill this criterion is with verification feedback, explicitly indicating if a step is right or wrong. Instead, grounded feedback supports students in evaluating their own work by supplying a second, external representation that makes relevant features more salient. Considered within Ainsworth's DeFT framework (Ainsworth 2006), this use of multiple representations is intended to facilitate correct interpretation of the novel representation (that is, constrain by familiarity). Returning to the number line example from the introduction, say a student is asked for a fraction that is smaller than $1/10$, and responds with $1/4$. Explicit verification feedback would indicate that $1/4$ is incorrect, and text elaboration could explain " $1/4$ is greater than $1/10$." However, if both fractions were plotted on a number line, an external, graphical representation, the student is likely to come to both of these conclusions on her own.

The second criterion is that *the feedback must be intrinsic to the domain and reflect the students' inputs with a linked, external representation*. This criterion draws from Dugdale's *intrinsic models* (Dugdale 1992), and distinguishes between verification feedback and feedback from a second, external representation. With verification feedback, the mapping from a student's input to the feedback depends on the problem being asked – an answer that is correct for one problem may be incorrect for another. With intrinsic feedback, the mapping from input to feedback is consistent, since the feedback simply reflects the student's input in another external representation, and the mapping between the input and feedback representations does not change based on the problem. For example, one intrinsic representation of an equation with two variables is a graph. With grounded feedback, when the student changes the equation, the graph changes to

reflect that. By going back and forth between acting on the equation and seeing a response in the graph, the student may better understand how the two are connected (e.g., ‘Green Globes,’ Dugdale 1992).

The combination of the first and second criteria mean that the feedback representation cannot just be any intrinsic representation, but must be one that is more accessible than the novel representation. That is, when the student is deciding if their work is correct, or if an action reduces the magnitude of an error, those tasks must be easier with the feedback representation than with the novel representation. And, the task of deciding if an answer is correct or not must be easy with the feedback representation. Therefore, the feedback representation must be chosen carefully to match the prior knowledge of the student. With the pre-requisite prior knowledge, a student is likely to generate the correct reference values for interpreting the feedback and deciding what to try next. A consequence of these two criteria is that with grounded feedback, students have the tools to evaluate their own work, even without explicit verification feedback. The act of leveraging prior knowledge for self-evaluation may deepen conceptual knowledge and strengthen the links between students’ ideas. Ohlsson (1996) provides a theoretical basis for this learning mechanism: evaluating our work often activates knowledge that was not available when the mistake was made in the first place (that is why we often correct mistakes when we check our own work, even without new information). Grounded feedback may help by making that evaluation step more explicit and by providing information, which a novice can interpret, about why their action was in error (Powers 1973).

Intrinsic feedback will usually fulfill the third criterion also: that *the feedback affords inferences on errors*. The feedback conveys information about the nature of errors, not just that a particular action was incorrect. For example, the feedback may indicate the direction or magnitude of the error. It is not required that students actually make these inferences in order for this criterion to be met, just that they are available. To meet this criterion, deciding if an action reduces the magnitude of an error must be easier with the feedback representation than with the novel representation. Returning to Ainsworth’s DeFT framework (Ainsworth 2006), this criterion for the use of multiple representations is intended to help students construct deeper understanding by mapping between both representations. Students should be given multiple chances to correct their errors through cycles of attempts and feedback.

The last criterion relates to the way students interact with the feedback rather than the design of the feedback representation itself: *The only way to change the accessible representation is through acting on the novel representation - students do not change or manipulate the feedback representation directly*. While students are expected to use the feedback representation to evaluate their work, generating that work occurs with the novel representation. Requiring that inputs be in the novel representation ensures some level of student engagement with that representation and is intended to promote transfer to contexts where the second external representation is not available. This criterion is based on prior work on concrete and abstract representations. Students learning from direct manipulation of concrete representations often have difficulty transferring their knowledge to symbolic contexts (Resnick and Omanson 1987; Uttal et al. 2013). From a situated cognition perspective (Lave 1988), acting on concrete representations produces knowledge that is tied to their use and affordances and will thus be difficult to access without them. Another explanation for this minimal transfer is that while actions

on concrete representations may be analogous to actions on symbolic representations, the cognitive processes of one may not require or even involve the cognitive processes of the other (Sarama and Clements 2009). The differences in cognitive process required by acting in the different representations may be so great that students may not even realize what their concrete manipulations are intended to model - for example, students moving a token a set distance along a number line may not realize their actions are intended to model addition (Suh et al. 2005).

Darts: an Example of Grounded Feedback

Darts (Dugdale 1992) is one example of grounded feedback (Simcalc is another – see Roschelle et al. 2000). In Darts, students enter a number or expression to shoot darts, attempting to pop balloons on a number line (Fig. 2). The dart flies to that location on the number line and stays there, with the original numeric input beside it. If a dart touches a balloon, the balloon pops. This example illustrates the four characteristics of grounded feedback:

- 1) *Students can easily envision the feedback state that indicates a correct answer.* When a balloon pops, it is clear that the dart hit the balloon.
- 2) *The feedback is intrinsic to the domain and reflect the students' inputs with a linked, external representation.* The placement of the dart on the number line is governed by the underlying mathematics, and the same magnitude information is conveyed in both the symbolic and graphical representations. A dart will always land at its specified location on the number line, even if no balloon is there.
- 3) *The feedback affords inferences on errors.* When a dart does not hit a balloon, the feedback gives information on the nature of the error. By comparing the dart's location to that of the target balloon, the student can tell if a larger or smaller

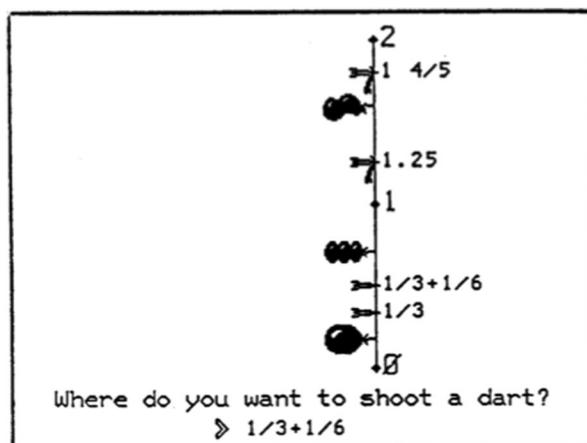


Fig. 2 Darts provides grounded feedback on students' numeric input. Here, the student has already popped balloons at $1\frac{4}{5}$ and 1.25. Reprinted from *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches* (p. 23), J. Larkin and R. Chabay (Eds.), 1992, Hillsdale, New Jersey: Lawrence Erlbaum Associates. Copyright 1992 by Lawrence Erlbaum Associates. Reprinted with permission

number is needed. Further, the feedback representation facilitates a rich set of inferences, for example, the student could infer that “ $1/3 + 1/6$ ” is about halfway between $1/3$ and the balloon target above, and thus infer that “ $1/3 + 2/6$ ” might be a good next entry. Since the goal is for students to understand how the symbolic numbers represent magnitude, each example of mapping between the numbers and their position on the number line is an opportunity for learning, even if the darts do not hit their target balloons.

- 4) *Students do not directly manipulate the feedback representation* Students cannot pop the balloons directly, and cannot interact directly with the number line. The only way to pop the balloons or to see the numeric value for a specific part of the number line is to write a number or equation. That is, students can only act directly on the novel representation.

Darts uses an intrinsic representation, the number line, to help students connect their understanding of magnitude to novel numeric symbols, including fractions, decimals, and expressions. A key feature of the number line for the intended audience of Darts is that it is more accessible than the numeric symbols. A student comparing two guesses on the number line is more likely to know which one is closer to its target than if the values for the guesses and target were presented symbolically. However, intrinsic representations are not all equally accessible. If students' numeric, base-10 inputs were reflected in hexadecimal, that would not make it easier for most elementary school students to decide which of two guesses is closer to the target. Therefore, hexadecimal feedback would not be grounded. Meeting the criteria for grounded feedback requires a match between the feedback representation and the student's prior knowledge: while the feedback will be grounded for students who have the pre-requisite knowledge, it will not be grounded for students without it. However, this does not necessarily mean that when feedback is grounded it will lead to learning, or better learning than alternative forms of feedback.

Theoretical Context: Prior Feedback Research

Past reviews of instructional feedback (Hattie and Gan 2011; Mory 2004; Narciss 2008; Shute 2008) have distinguished different kinds of feedback and, in some cases, indicated differential impact on student learning. Shute (2008) defines formative feedback as “information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning” (Shute 2008, p. 154). As such, grounded feedback is one type of formative feedback. Grounded feedback follows Shute's (2008) recommendations that feedback be objective, focused on the task (not the learner), and presented visually (instead of using text alone). Further, Shute (2008) indicates that elaborative feedback, which provides cues and information, may be especially helpful for low-ability students, while verification feedback, which indicates if a response is correct or not, may be more helpful for high-ability students. Grounded feedback is elaborative, as it shows the nature of students' errors, and it can also be used for verification. However, Shute's review focused on explicit feedback, such as checkmarks for verification and hints or prompts for elaboration (Shute 2008). In contrast, the facilitation, elaboration, and verification in grounded feedback are all implicit – they

require some reasoning on the part of the student. While the student works with the novel representation, grounded feedback reflects those actions in a linked representation that is more accessible. That is, the student can easily connect the feedback representation to their prior knowledge. When the student does so, the feedback representation may act as a bridge to link the students' prior knowledge to the novel representation. Grounded feedback is intended to strengthen students' mental connections between their prior knowledge, the new content they are learning, and the novel and accessible representations – a type of knowledge termed “integrated-concrete” (Sarama and Clements 2009).

Grounded feedback is consistent with Hattie and Gan's emphasis on “mindfulness” (Hattie and Gan 2011) by being a non-evaluative consequence of a student's actions, which the student must actively interpret before deciding if it indicates a correct answer. Grounded feedback also responds to Mory's call for further research on feedback for helping students correct errors (Mory 2004). By encouraging students to engage meaningfully with their errors (without giving them the correct answers), grounded feedback can also be seen as a form of informative tutoring feedback (Narciss 2008). The criteria for grounded feedback and the proposed mechanisms for its benefits fit within the interactive, two-feedback-loop (ITFL), a framework proposed by Narciss (2008). In this framework, an external feedback loop is managed by the tutoring system, teacher, or learning environment, while the internal loop takes place in the mind of the student. The internal loop processes the external feedback in conjunction with internal reference values to produce internal feedback and then generate the learner's next step (Narciss 2008). Grounded feedback is intended to promote a specific type of internal loop, where the learner draws on relevant prior knowledge that she would not have been able to apply to the novel representation directly, but that she can easily apply to the accessible representation.

Theoretical Context: Prior Work on Grounding in Argumentation and Communication

We use the term *grounded* because the feedback is grounded both in a different representation and in the student's prior knowledge. Darts dynamically translates a student's intention (estimate a balloon's location with a number or expression) to a representation (the number line) that may help the student see if that action matched the original intention (is the target balloon located at that spot?). It lets the system ask “is this what you mean?” perhaps giving pause when the feedback is not what the student expected. While the term *grounded* should not be confused with Toulmin's (2003) *grounds* in argumentation or Clark and Brennan's (1991) *grounding* in communication, they have some commonalities. *Grounds* are “those things which have to be specified in reply to the question, ‘How do you know?’ before an assertion need be accepted as justified” (Toulmin 2003, p. 223). Considering a student's interaction with a tutoring system within Toulmin's framework of argumentation, we may think of the student's initial answer as a claim (e.g., the large balloon is located at $1/3$). Grounded feedback provides the *grounds* that help justify or refute the claim (e.g., representing the number $1/3$ on the same number line as the large balloon provides grounds for rejecting the claim that they are located at the same spot). *Grounding* in communication is the process of establishing “mutual knowledge, mutual beliefs, and mutual assumptions”

(Clark and Brennan 1991, p. 127). Considering interactive tutoring as a form of communication between the learner and the system, the novel representation may be a medium without “mutual knowledge” between the learner and the system – the learner may think that $1/3$ indicates a different magnitude than the system does. Grounding in communication establishes what the listener understands the speaker to be saying, perhaps indicating that the speaker should clarify if the message being received does not match the message that was intended. Likewise, grounded feedback shows the meaning of the student’s input in a representation that allows the learner to tell if the input accurately conveyed her intentions or not.

Another commonality between these three concepts is that successfully creating grounded feedback, or grounds for an argument, or grounding in communication, depends not only on the actions of the system, the debater, or the speaker, but also on the mental state of the student or the listener. Grounds in an argument may not be accepted by the listener – they may be interpreted as claims which require further evidence. Grounding in communication may fail if a clarification is insufficient. While an important aspect of grounded feedback is that the second representation be easy for the learners to interpret, they may not have the prior knowledge to do so. Darts will only be grounded for students who understand the number line.

Contrasting Grounded Feedback to Related Instructional Approaches

Table 1 contrasts grounded feedback with explicit verification feedback, non-linked representations, linked representations where the accessible representation is manipulated directly, and linked representations that do not make relevant features more salient. The columns in Table 1 indicate which of the four criteria of grounded feedback are present in the other feedback types. The goal of this section is to highlight the

Table 1 Comparing grounded feedback to other instructional approaches

Systems	Grounded Feedback Features			
	Students can easily envision the feedback state that indicates a correct answer.	Feedback is intrinsic, and reflects student inputs.	Feedback affords inferences on errors.	Students act on the novel representation, not the accessible one.
Grounded Feedback	Yes	Yes	Yes	Yes
Cognitive Tutor (explicit verification)	Yes	No	No	No second representation
Chemistry models; QUADRATIC (non-linked representation)	Yes	No	No	Yes
Fractions Applet (accessible representation is manipulated directly)	Yes	Yes	Yes	No
RelLab	No	Yes	No	Yes

features of grounded feedback through contrast. The contrasted systems were selected as examples of real-world designs that differ from grounded feedback on at least one criterion. Further, these systems together show examples of the presence and absence of each criteria for grounded feedback. In several cases we hypothesize that grounded feedback may be more beneficial than the contrasted approach. However, in all of these comparisons, there is little or no empirical evidence for the superiority of one approach or the other. One goal of this paper is to call for experiments that would provide such evidence.

Verbal Feedback and Explicit Verification in Tutoring Systems

How is grounded feedback different from explicit verification? Consider the feedback provided by the Algebra Cognitive Tutor (Koedinger and Alevan 2007) shown in Fig. 3.

The tutor marks incorrect inputs ($.13t$ in the second column) and prevents students from progressing in the problem until that error is fixed. For recognized common errors, the system provides text feedback. Here, the tutor explains that the student's expression for the cost of t minutes of phone calls, $.13t$, does not include the base charge of \$14.95 per month. Explicit verification feedback meets the first criterion for grounded feedback, since students can easily tell if their answers are correct (they turn green). Explicit verification does not meet the second criterion since it does not provide an intrinsic representation of the students' inputs. It does not meet the third criterion since it does not afford inferences on errors – the feedback is the same for a large error as it is for a near miss. It does meet the fourth criterion since students act on the novel representation.

The key differences between grounded feedback and the Algebra Cognitive Tutor feedback is that explicit verification is not intrinsic to the domain and thus does not promote the same degree of inferences on the nature of the students' errors. Grounded feedback is intended to support conceptual reasoning, even in drill-and-practice environments. Figure 4 shows one possible redesign to make this tutor grounded. First, before constructing the expression, students calculate the cost of the current cell phone plan for various numbers of minutes (it is actually easier for novices to provide a

My current cell phone company charges me \$14.95 per month for service and \$.13 per minute. PPS Cellular Phone Company has offered me \$15.00 worth of free calls a month if I switch, but the charge is \$.39 per minute.

1. How many minutes of calls can I get from PPS Cellular Phone Company for \$50? What is the cost from my current company for that number of minutes?

Quantity Name	Time	Current cost	PPS cost
Unit	minutes	\$	\$
Expression	t	$.13t$	
Question 1			

The cost from my current company increases by 0.13 each minute, but remember that it starts at 14.95 dollars.

Fig. 3 Sample work and feedback with the algebra cognitive tutor. The student's symbolization $.13t$ is immediately marked as wrong, without giving the student the opportunity to evaluate his own work

Quantity Name	Time	Current cost (calculated)	Current cost (expression)	Expression for your current cost:
Unit	minutes	\$	\$	
	0	14.95	0	
	1	15.08	.13	
	2	15.21	.26	
	10	16.25	1.30	

Fig. 4 Proposed implementation for grounded feedback in an algebra tutor, showing the same student error as Fig. 3. First, the student calculates the charges for various numbers of minutes. After inputting an expression for the current cost, the tutor generates the costs for the given numbers of minutes. This allows the student to evaluate her own work, and to see what costs are generated by incorrect expressions

numerical answer to a story problem than it is to construct the symbolic expression that yields that answer; see Heffernan and Koedinger 1998). To guard against slips, students could get explicit verification on the calculated costs. After finding the correct costs, students propose an expression, which the tutor evaluates for each of the given number of minutes. This allows the students to judge the correctness of the expression by comparing its results to the correct cost that they just calculated. In the grounded feedback version, students can see what the costs would be if the charges were only 13 cents per minute. By comparing the values derived from the expression to the ones they calculated, the students are likely to determine for themselves if they have made an error. Upon considering the zero-minute row, this student is likely to see for himself which part of the expression he forgot.

Intelligent tutoring systems use immediate, explicit verification feedback to reduce the unproductive cognitive load that comes from floundering. However, as Fig. 4 shows, grounded feedback can be implemented in a way that is compatible with cognitive load theory. Since grounded feedback is calibrated to students' prior knowledge, the tasks and decisions students are given are ones for which the students have a high probability of success. This is in line with cognitive load theory design principles, which propose that students should perform problem-solving steps (including evaluating their work) if they are likely to do so efficiently (Kalyuga et al. 2001). Experiments comparing grounded to explicit verification feedback have found benefits for grounded feedback (see section 'Evidence on the Effectiveness of Grounded Feedback'). Comparisons of grounded feedback and explicit verification could answer several research questions, including: What aspects of the domain, students' prior knowledge, and students' meta-cognitive and self-regulation skills determine when one type of feedback is better than the other? What information do students extract from the grounded feedback that they would not infer from explicit verification? Does grounded feedback promote more connections between students' prior knowledge and the target knowledge?

Non-Linked Representations

We discuss two systems that use non-linked representations, as they appear very similar to grounded feedback and the differences are subtle. Padalkar and Hegarty's chemistry instruction (Padalkar and Hegarty 2012) uses physical models as the more-accessible

representation (similar to Izsak 2000), and the QUADRATIC tutor (Wood and Wood 1999) is entirely on the computer. In both cases, students work primarily in the less-accessible representation and use the more-accessible representation to check their work. Unlike grounded feedback, the two representations are not linked. Instead of showing the student's current work, the accessible representation only shows the correct answer.

In Padalkar and Hegarty's chemistry instruction (Padalkar and Hegarty 2012), students were given a diagram of a molecule and were asked to draw the same molecule with another type of diagram. Students were also given a three-dimensional ball-and-stick model (the non-linked representation), which they could compare against to check their work: if the components of the student's hand-drawn diagram could not be mapped to the ball-and-stick model, the student would know that the drawn diagram was incorrect. Providing the ball-and-stick model as feedback meets the first criterion of grounded feedback: students can tell that their work is correct when all elements of the hand-drawn model can be mapped to the ball-and-stick model. While the ball-and-stick model is an intrinsic representation, it does not meet the second criterion because it does not update to reflect changes in students' hand-drawn models. It does not meet the third criterion because it does not distinguish between different types of errors. It does meet the fourth criterion because students act on the novel representation (the hand-drawn diagram). Interestingly, students generally did not spontaneously use the non-linked representation, often because they were over confident in their answers (Padalkar and Hegarty 2012). In a controlled experiment with the models available at every stage, students were given a pretest, explicit instruction in how to use the models to check their work, and a post-test. Students who were told how to use the non-linked representation ended up using the models more at post-test and improving more from pre- to post-test than a comparison group that had access to the models but no explicit instruction in how to use them (Padalkar and Hegarty 2012). This suggests that while non-linked representations can be helpful for self-evaluation and learning, students may not use them, and therefore may not notice discrepancies between their work and the non-linked representation.

One way to help draw students' attention to the more-accessible representation is to automatically map the students' answers onto the non-linked representation. For example, in the QUADRATIC tutor (Wood and Wood 1999), students expand quadratic expressions such as $(x + n)^2$. As students form symbolic expressions, the tutor maps correct terms onto a geometric model (a partitioned square with sides of length $x + n$; see Fig. 5). QUADRATIC meets the first criterion of grounded feedback: students can tell that their work is correct when all terms of their equation are mapped to the geometric model. While the geometric model is an intrinsic representation, it does not meet the second criterion because it does not appear to update to reflect incorrect terms in students' equations. It does not meet the third criterion because it does not distinguish between different types of errors. It does meet the fourth criterion because students act on the novel representation (the equation). Making QUADRATIC grounded requires only one change: that the geometric model reflects all terms in the student's equation, including incorrect terms.

It is not clear when grounded feedback or non-linked representations are better for learning, and we are unaware of any experiments that compare the two. We offer two competing perspectives: (1) The non-linked approach is better because it encourages

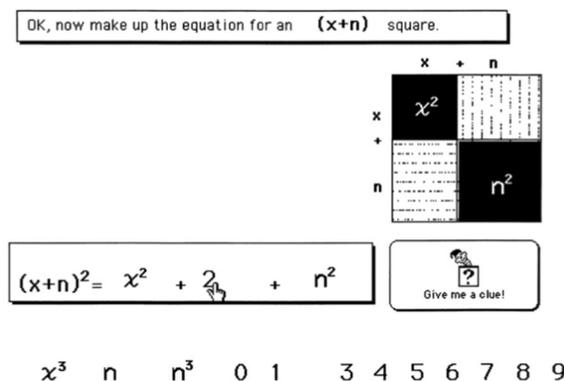


Fig. 5 Screenshot from the QUADRATIC tutor. Students' correct numeric inputs are mapped to a geometric model. However, unlike grounded feedback, the geometric model does not appear to reflect incorrect inputs. Reprinted from "Help seeking, learning and contingent tutoring," by H. Wood and D. Wood 1999, *Computers & Education*, 33, (p. 157). Copyright 1999 by Elsevier. Reprinted with permission

students to actively integrate the two representations; (2) The grounded approach is better because it helps students diagnose their own errors in mapping between the two representations. To learn effectively from multiple representations, students must integrate them and know how to map between them (Ainsworth et al. 2002). In grounded feedback systems, a computer does this mapping by generating a reflection of the student's actions in the feedback representation. In non-linked representations where a mapping is not provided, the student must do this mapping to evaluate her work. It is possible that this active student integration is a key ingredient for student learning with these systems, a view supported by evidence showing that students learn more when they integrate static representations rather than view pre-integrated ones (Bodemer et al. 2005). If active student integration improves learning, students would learn more from a non-linked system where the mapping is not provided than from either a non-linked system where mapping is provided or from grounded feedback.

On the other hand, as discussed above, students may not spontaneously map between the two representations. Even when students do map between the two representations on their own, they may do so incorrectly. For example, an algebra student trying to convey "40 subtracted from 800" may write "40 - 800," not realizing that the minus sign can only mean "subtract" and never "subtracted from" (Koedinger, personal communication). In that case, an accessible representation showing 760 would not help the student realize that his work was incorrect. Further, the novel representation itself may trigger misconceptions. Roschelle et al. (2000) discuss students' misconceptions around an elevator simulation, where an elevator goes up and down at various speeds, represented by a piecewise linear function of velocity. While interpreting these graphs, students often confuse "going down" with "slowing down," thinking that a decrease in velocity represents the elevator moving downward, when it actually represents the elevator moving upward, but more slowly (Roschelle et al. 2000, p. 18). If the student is left to interpret these less-accessible representations on her own, she may do so incorrectly and not realize it. If students are likely to map incorrectly or not at all, grounded feedback systems may be more beneficial than non-linked systems. Comparisons of grounded feedback and non-linked representations could answer

several research questions, including: When does active integration of two representations lead to better learning than grounded feedback? Are students more likely to detect their errors with grounded feedback or non-linked representations? Do students learn better self-correction skills from systems that differentiate between errors (grounded feedback) or not (non-linked representations)?

Fraction Equivalence Applets: a Linked Representation, Opposite Direction

The Equivalent Fractions applet from the National Council of Teachers of Mathematics 2014 (Fig. 6; illuminations.nctm.org) is one example of a linked representation that connects an accessible representation (fraction rectangles) to a novel representation (fraction symbols). The key difference between this feedback and grounded feedback is the direction of the link: Grounded feedback uses the less-accessible representation as input and the more-accessible one as feedback, while this applet does the reverse. The equivalence applet presents one fraction and asks the student to generate two equivalent fractions with different denominators. Each proper fraction n/d is represented in three ways: as a fraction rectangle with d equal pieces, n of which are colored in; as a location on a zero-to-one number line; and as a symbolic fraction. Students generate equivalent fractions by moving the horizontal and vertical sliders next to each fraction rectangle to create equal divisions, and then click on the divisions to color them in. The corresponding symbolic fractions and number line locations are updated to reflect the students'

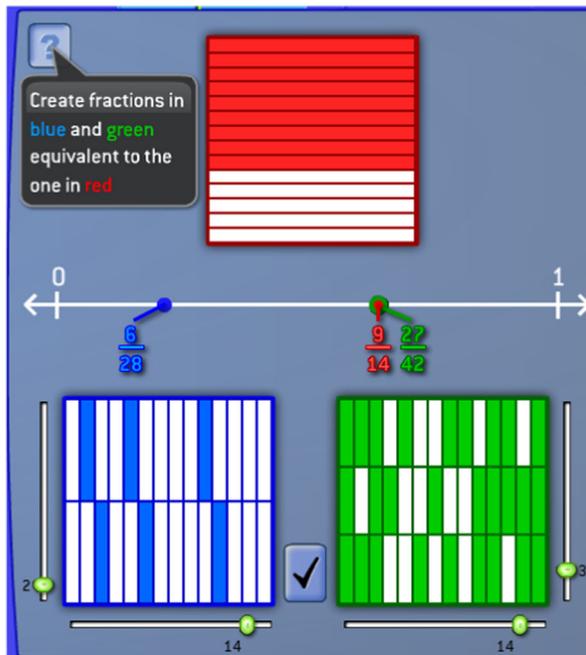


Fig. 6 Students input fractions by using the sliders to generate equal-sized pieces and then click on each piece to color it in. The symbolic and number line representations are dynamically linked to the rectangles. The feedback is not grounded since students manipulate the more-accessible representation and get feedback in the less-accessible representation. Reprinted from Illuminations (illuminations.nctm.org). Copyright 2014 by the National Council of Teachers of Mathematics. Reprinted with permission

manipulation of the rectangles. To confirm that the fractions are equivalent, the student can press the check button. The applet meets the first criterion of grounded feedback: students can tell that their work is correct when all three points overlap on the number line. (However, it is unlikely that students would be able to envision the symbolic fractions that indicate equivalent fractions – if they could, this lesson would not be necessary.) It also meets the second criterion: the number line and the numeric fractions are both intrinsic representations for the fraction rectangles. It also meets the third criterion, affording inferences on errors. By comparing the relative positions of the points on the number line, a student can tell if her answer was too big or too small, close or far. The applet does not meet the fourth criterion. Unlike grounded feedback, in this applet students directly manipulate the more-accessible rectangle representation, while a less-accessible representation (the symbolic fractions) are provided as feedback. Since students are not acting directly on the novel representation, they are unlikely to be practicing the cognitive steps that will transfer to a symbols-only context. While dividing a rectangle with one of the sliders is *analogous* to multiplying the current denominator by some number, that action itself does not require the student to think about multiplication. Though the applet may help students solidify equivalence concepts, it does not give students practice with symbolic procedures, and thus is unlikely to efficiently lead to robust learning of the procedures or robust understanding of how the procedures and concepts are connected.

One way to make this activity grounded is presented in Fig. 7. Students input symbolic numbers in the areas marked with black borders. The equations encourage students to multiply, but the interface does not require it. If students choose to multiply the denominator, the original horizontal divisions are overlaid with vertical divisions. If the student chooses not to multiply, the rectangle shows only horizontal divisions. The student-entered numerator determines how many pieces are colored, and those pieces are colored consecutively. Since the rectangles are aligned, students can compare the magnitudes without the number line. Alignment also should facilitate comparison of

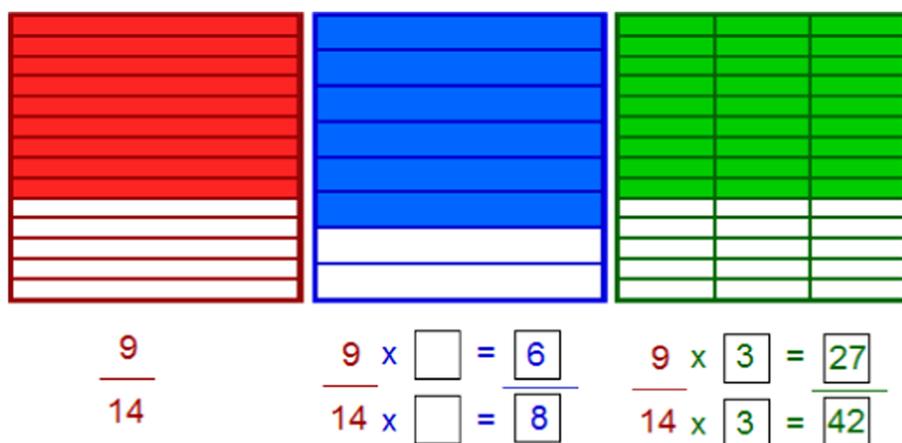


Fig. 7 Proposed re-design of equivalence applet that uses grounded feedback. Students enter symbolic numbers in the black-bordered input areas, and see the rectangle representation of the fractions as feedback. The equations encourage students to multiply, and to think about multiplication when they interpret the feedback

denominators (e.g., eighths do not line up with fourteenths). Although we hypothesize that grounded feedback will be more beneficial than accessible-to-less-accessible linked representations, we hasten to add that we have not found experiments comparing these two designs. Linked representations from the more-accessible one to the novel one is a popular design choice (e.g., PhET, <https://phet.colorado.edu/> (University of Colorado, Boulder, 2014), Shodor 2014, <http://www.shodor.org/interactivate/>, and the National Library of Virtual Manipulatives (Utah State University 2014), <http://nlvm.usu.edu/>), and evidence is needed on whether this is the most beneficial choice for learning. Specifically, comparisons of grounded feedback and linked representations from the more-accessible one to the novel one should investigate which promotes better transfer to contexts where the student is only working with the novel representation. Additionally, eye-tracking comparisons should investigate if students pay more attention to the novel representation when it is the input representation or the feedback representation.

RelLab: Feedback that is Situational, but not Grounded

Situational feedback draws on theories of a *problem model* and a *situation model* when students encounter story problems. While the *problem model* “represents the mathematical structures needed to solve the problem” (Nathan 1998, p. 139), the *situation model* “draws on the reader’s prior knowledge of events and semantic knowledge” (Nathan 1998, p. 139). Experts draw on both models. When novices rely on the problem model to the exclusion of the situation model, they may generate nonsensical answers (such as answering “31 remainder 12” to a question asking how many 36-passenger buses are needed to transport 1128 soldiers; Schoenfeld 1988, p. 6). Situational feedback aims to help students connect the mathematics in the problem model to the situation model. While some implementations of situational feedback are grounded, some are not.

The RelLab simulation environment (Horwitz and Barowy 1994) is an example of situational feedback that is not grounded. RelLab was designed to teach physics concepts to high school students, and performs simulations of events at normal and relativistic speeds. RelLab includes a physics simulator that animates scenarios that the students input. For example, one question describes a scenario where one car is following another, going at the same speed, both with identical 70-mile-per-hour baseball pitching machines mounted to the roof and facing the other car. The question asks if each driver would see their baseball as moving faster, the same speed, or slower than the other character’s, and what a stationary observer on the sidewalk would see. To answer the question, students interpret the scenario and input parameters to the simulator, and then watch the animation play out. For the task of selecting the input parameters, this environment may appear to meet all of the criteria of grounded feedback: (1) by seeing the animation play out, students can tell if their inputs correctly model the physics scenario; (2) the animation is an intrinsic representation of the students’ inputs; (3) the animations afford inferences on errors since different types of errors will result in different animations; and (4) students do not directly manipulate the animation – the only way to control it is through the symbolic parameters that define the scenario. However, students’ interactions with the system revealed that it was not grounded for them. Seeing the animations did not make it easy for students to tell when they had chosen incorrect parameters for the situation. A common mistake students made was to select an incorrect reference frame for the baseballs, defining

them as moving 70 miles per hour relative to the ground rather than relative to the car that held the machine. However, watching the animations did not alert students to their mistake, because they did not have enough prior knowledge of physics to envision what the correct animation would look like. When the target content involves conceptual change, students are unlikely to correctly envision the goal state, meaning that they cannot effectively use their expectations to evaluate the situational feedback. In other words, for some target content, the knowledge required to interpret the grounded feedback may be the same as the ultimate learning goals of the lesson.

RelLab illustrates how situational feedback is different from grounded feedback. Situational feedback is a property of the learning environment, but grounded feedback is a property of the match between the learning environment and the student's prior knowledge. Our point is not to say that RelLab is a poor learning environment (on the contrary, it was successful in prompting discussions and helping students understand physics), but rather that story problems and feedback given in a situated context does not necessarily make it easier for students to tell if their work is correct. Conversely, Darts is an example of a system that is grounded but not situational. While Darts does involve grounded feedback, it does not involve a story problem or require the student to generate a "situation model" based on semantic relationships and prior knowledge of events. Therefore, we argue that grounded feedback and situational feedback may overlap, but one is not necessary or sufficient for the other.

Comparing Feedback Types

One way to compare grounded feedback, explicit verification, non-linked representations, and linked representations from accessible to novel would be to implement four variations of the same system. The section "Fraction Equivalence Applets: A Linked Representation, Opposite Direction" gives examples of two of the four designs (grounded feedback and linked representations from accessible to novel). To implement the applet with a non-linked representation, students could be given the correct answer in the fraction rectangle format (but they would still need to produce their answer in the numeric format). To implement the applet only with explicit verification, students would only be given numeric representations of the fractions, and their answers would immediately be marked correct or incorrect.

Empirical Evidence on the Effectiveness of Grounded Feedback

The papers in this section were selected because they describe experiments that use random assignment to compare a grounded feedback system to a control, and they measure learning with pre- and post-tests. In addition to the relevant papers that were already known to us, we searched the databases Google Scholar, ERIC (peer-reviewed results only), and PsycInfo with the keywords "Grounded Feedback," "Situational Feedback," and "Linked Representation" (with "Linked Representation Education" for Google Scholar). For Google Scholar and ERIC, we considered the first 40–60 results for "Grounded Feedback" and "Situational Feedback" and the first 70 results for "Linked Representation." For PsychInfo, we considered all search results within relevant index terms and classifications, such as "Learning Environment" or "Educational

Psychology.” This search yielded one additional paper that met the inclusion criteria (using random assignment to compare a grounded feedback system to a control, with pre- and post-tests to measure learning).

ANIMATE: Comparing Grounded Feedback to Error Messages

Nathan’s ANIMATE system (1998) is an implementation of situational feedback. The ANIMATE tutoring system teaches students how to model a story problem with algebra equations. Students set up equations, which drive animations, which the student can then compare to the situation in the story. A sample problem: a train leaves its station going 75 miles per hour. A helicopter leaves from the same station two hours later, going 300 miles per hour, to warn the train that there is a broken bridge 60 miles ahead. Can the helicopter catch up with the train in time? Figure 8 shows a sequence of example student work and feedback for this problem. Figure 8a shows the student’s expectation that after an hour, the train will have traveled 75 miles and the helicopter will not have left the city yet. Figure 8b shows the system of equations the student has entered to model the story problem: $D1 = D2$ (both vehicles have traveled the same distance once the helicopter catches up with the train); $D1 = 75 * T1$ (the train’s distance is its speed multiplied by its travel time); $D2 = 300 * T2$ (likewise for the copter). $T1 = T2 - 2$ relates the amount of time that the two vehicles have been traveling, demonstrating a common misconception. The student tried to model that the copter leaves two hours after the train, perhaps thinking if the train left at 9 am, the helicopter would have left at 11 am, and $9 = 11 - 2$. However, the equation requires that $T1$ and $T2$ represent the amount of time each vehicle has been traveling, not the clock time when they left. Therefore, the correct equation is $T1 = T2 + 2$ since the train travels for two hours more than the helicopter. These entered equations drive animations of the train and the helicopter. Figure 8c shows the animation for the positions of the train and helicopter after an hour. Unlike the student’s expected outcome in Fig. 8a, c shows that the helicopter travelled 300 miles and the train stayed at the station. In this

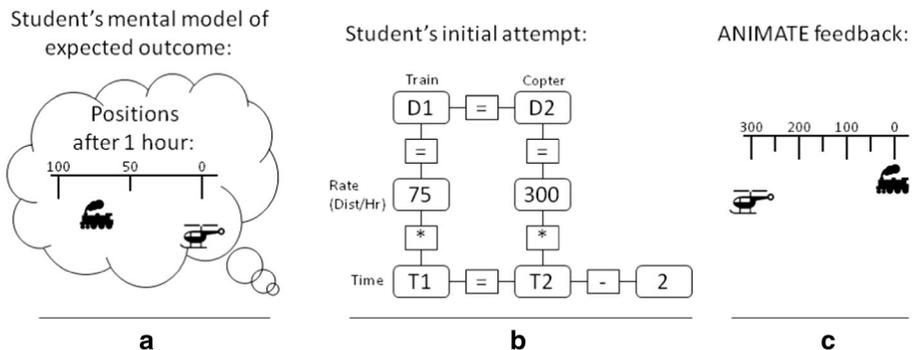


Fig. 8 a The student’s expectation that after one hour, the train will have gone 75 miles and the helicopter will not have left. b The student’s inputted equations for modeling the story problem. c ANIMATE’s feedback, based on the entered equations, does not match the student’s expectations. The feedback supports qualitative reasoning, while the input format supports algebraic symbolization. Adapted from “Knowledge and Situational Feedback in a Learning Environment for Algebra Story Problem Solving,” by M. Nathan 1998, *Interactive Learning Environments*, 5, (p. 141). Copyright 1998 by Taylor & Francis Ltd. (www.tandfonline.com). Adapted with permission

example, the animation would show the chase helicopter leaving *before* the train, which does not match the problem.

The ANIMATE system is both situational and grounded. ANIMATE is not an intelligent tutoring system: it has no student model, does not mark answers as correct or incorrect, does not provide text hints, and does not ensure that students attain the correct answer before moving on to the next problem. Instead, its key feature is providing student-meaningful situational feedback. ANIMATE is one of the few grounded feedback systems that has been compared to a control, with learning measured with pre- and post-tests outside the tutors (Nathan 1998). Instead of running simulations, the control tutor gave a sequence of three pop-up hints when students made errors (e.g., at the error depicted in Fig. 8, the first hint reads “It is common to over-generalize ‘later than’ to mean minus. Please check your current work.”) An experiment with 31 college students using a pretest-intervention-posttest design showed that while both groups improved in modeling story problems from pre- to post-test, the situational feedback group improved more. The tests included one problem of each type: travel (example show in Fig. 8); investment (e.g., \$750 is invested at an interest rate of 5%, compounded annually. How much is in the account at the end of the second year?); and work (e.g., Tom can paint the entire fence in two hours while it takes Huck four hours. If Tom arrives one hour late from fishing, how long will it take the two boys to complete the job?). Separate ANCOVAs for each problem type with pretest performance and total SAT (the standardized test) as covariates and treatment as a between-subjects factor showed a significant difference for treatment ($p < .05$) on travel and investment problems, with .76 for the standardized gain for the situational feedback condition on both types and .57 and .43 for the control, respectively. The ANCOVA for work problems did not show a significant difference between conditions.

While Nathan's 1998 study shows strong overall benefits for situational/grounded feedback (relative to the pop-up text hints), student learning was not significantly different on work problems. Work problems differed from the other types in that students were given a whole number in the problem statement (Tom can paint 2 fences per hour) but they needed to use the reciprocal in the equation (Tom needs 1/2 an hour per fence). ANIMATE students could see that using the original number was incorrect, but did not know what to try next. This finding suggests that grounded feedback may only be more beneficial than explicit verification feedback or pop-up text hints when students are able to both recognize that they have made an error and make useful inferences on those errors to guide them on what to try next. Alternatively, a system that provides both grounded feedback and text hints (when students are stuck) may be more powerful than either type of support on its own. Mathan and Koedinger's Excel tutor (2005) is one such system.

Excel Tutor: Comparing Grounded Feedback to Explicit Verification

Mathan and Koedinger's 2005 Excel tutor teaches students how to write spreadsheet formulas with absolute and relative cell references. In the grounded feedback version of the tutor, Excel evaluates each formula that the student enters. From Excel's feedback (providing the calculated values for each formula), the student can determine if the original formula was correct. For example, one problem asks students to calculate the interest owed on a loan of \$10,000, at various interest rates (Fig. 9). The problem

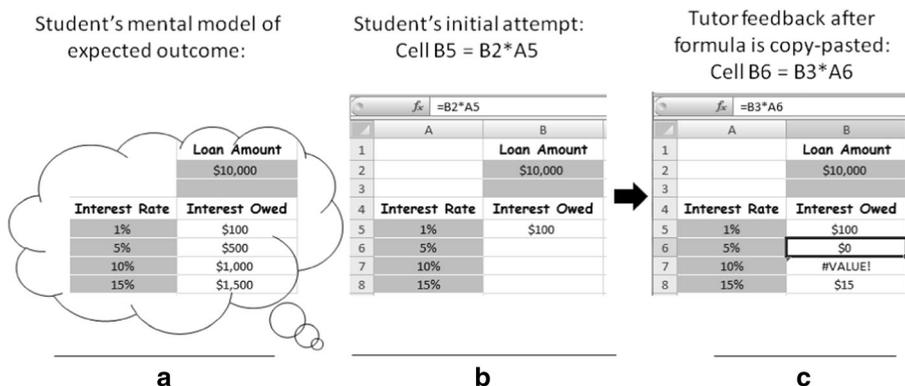


Fig. 9 **a** The student mentally calculates the interest rates. **b** The student multiplies the first interest rate (cell B2) by the loan amount (cell A5), yielding the correct interest for cell B5. **c** When that formula is copied and pasted in cells B6–B8, it multiplies the interest rates by cells under the loan amount. This result does not match the student's expectations. Adapted from *Educational Psychologist*, 40, S. Mathan & K. Koedinger, "Fostering the intelligent novice: Learning from errors with metacognitive tutoring," p. 263, Copyright (2005), with permission from IOS Press. This publication is available at IOS Press through http://dx.doi.org/10.1207/s15326985sep4004_7

content was designed so that students would likely be able to calculate each of the interests owed, or at least recognize clearly incorrect values (Fig. 9a). Figure 9b shows the student-entered formula “=B2*A5” in the cell B5 (entry shown at the top), which multiplies the loan amount by the 1% interest rate. Excel responds with “\$100” in B5 and this value matches the student's expectations. However, when the student copies the formula from B5 and pastes it in cells B6 through B8 (Fig. 9c), Excel's values do not match the student's expectations. The interest owed at the 5% rate is shown to be \$0. Upon inspecting the formula for that erroneous cell (top of Fig. 9c), the student is intended to see that the interest rate is not being multiplied by the loan amount, but by the cell directly underneath. Excel multiplied the wrong cells because the student used a relative instead of an absolute reference (the correct formula for B5 is “=B\$2*A5”). If the student cannot fix the error independently, the tutor provides step-by-step guidance. The grounded feedback tutor (which Mathan and Koedinger called an “Intelligent novice model spreadsheet tutor”) was compared to a version that gave explicit interactive support as soon as students entered an incorrect formula. In this control condition, students had to generate the correct formula before pasting it into multiple cells. In both conditions, the tutor offered text hints if students needed them. In the intelligent novice condition, students could (1) see how Excel responded to incorrect formulas; and (2) try to recognize and correct their own errors before the tutor jumped in.

An experiment with 49 adult job seekers using a pretest-intervention-posttest design showed that, like the ANIMATE experiment, while both groups improved from pre- to posttest, the grounded feedback group improved more (Mathan and Koedinger 2003, 2005). Students in the grounded feedback condition showed significantly better learning from pretest to posttest on all of their measures, with substantial effect sizes (across all treatment-to-control comparisons) for problem solving (effect size: .50), conceptual understanding (effect size: .59), transfer (effect size .43), and retention (effect size: .33). These strong results show the additive benefit of grounded feedback in a learning environment that already provides text hints.

Fractions Tutor: Comparing Grounded Feedback to Worked Examples

While the experiments discussed above showed robust learning benefits for grounded feedback, other related work has found no differences in learning or inconclusive results. One study on fractions found that linked representations did not outperform worked examples (in the context of a larger investigation of learning with multiple representations; Rau et al. 2012). The linking was similar to grounded feedback in that students worked with number lines (a less-accessible representation) and got feedback on their actions with fraction rectangles or circles. However, the linking was not a consistent implementation of grounded feedback: in some cases the second representation was static and only depicted the correct answer, and in some cases the dynamic representations could only reflect a subset of students' inputs on the number line. This study suggests that worked examples may be more effective than grounded feedback when three representations are involved, but due to the inconsistent implementation of grounded feedback in this system the results remain inconclusive.

Linear Relationships: Comparing Grounded Feedback to Linked Representations

Another study investigating several representations examined students' learning of linear relationships (Ozgun-Koca 2004). The representations included a video depicting linear movement of two objects (e.g., of two fish swimming past each other) and a table and graph showing the positions of the objects over time. In the fully-linked condition, students' selections on the video, table, or graph would highlight the corresponding data in the other representations. Students could also view an equation for the best fit line, with the line plotted on the graph. In the semi-linked condition, students could estimate the coefficients for the best fit line, and see their estimate plotted on the graph (however, students' selections of points in one representation did not cause highlighting in the other representations). For the purposes of evaluating grounded feedback, this study is confounded: the semi-linked condition provided grounded feedback for students' estimates of the coefficients for the line of best fit, while the fully-linked condition did not, but the two conditions also differed in how the other representations were linked to each other. This study found no differences in learning between the two conditions, though with 10 students in each condition the study may have been underpowered.

TRANSFORMER: Comparing Grounded Feedback to Explicit Verification

A study on algebraic transformations compared four conditions: grounded feedback, problem-level explicit verification, problem-level explicit verification with on-demand demonstration of transformation steps, and no feedback (Yerushalmy 1991). In each version of the tutor, students were given an algebraic expression and had to transform it into a different format (e.g., given $x(x-2)^3 + 7(x-2)$, change it into the format $Ax^2 + Bx + C$). In the grounded feedback condition, students were shown three graphs: one of the original expression, one of the student's current work, and one showing the difference between them. An experiment with 7th graders used a pretest-intervention-posttest design to measure learning outside the tutor; test data was reported for 17 students. All groups improved from pre- to posttest on the target content, without

significant differences in learning between the conditions. While this study is quite small and likely underpowered, Yerushalmy identified some qualitative pros and cons of the grounded feedback. First, it appeared that the feedback was indeed useful in helping students evaluate if their attempts were correct – compared to the no-feedback condition, grounded feedback students performed more steps per problem, did more relevant debugging, left fewer uncorrected errors, and ultimately solved more problems correctly during tutoring. For students with high prior knowledge, the graph feedback helped them locate which term was the source of their error. However, other students, especially those with low prior knowledge, reacted to the grounded feedback as problem level verification feedback or simply tried to eliminate the difference graph without looking for the source of the error, a form of gaming the system.

Fraction Addition Tutor: an Attempted Implementation of Grounded Feedback

Our final example in this area is a controlled experiment with a fraction addition tutor attempting to use fraction bars as grounded feedback (Wiese 2015). In the fraction bar tutor, students were shown fraction addition questions with the fractions represented symbolically and with fraction rectangles (Fig. 10). When the students converted the addends and inputted sums, the system would reflect those inputs as fraction rectangles, which could then be compared to the given fraction rectangles to verify equality. In a pilot study, students used the fraction bars effectively to decide if their work was correct. One student solving $2/8 + 3/8$ initially added the numerators and denominators, yielding $5/16$. The student looked surprised when the fraction bar feedback showed her answer to be much smaller than the multi-colored bar representing the sum. She then corrected her answer to $5/8$. In the controlled experiment, the fraction bar tutor provided explicit problem-level verification feedback, as students were not permitted to move on to the next problem until the current one was solved correctly. However, the tutor did not have explicit step-level verification feedback. The control condition did not include any rectangles, and instead provided immediate, explicit step-level verification, coloring inputs green if correct and red otherwise. Both tutors offered on-demand text hints. This study used a pretest, intervention, posttest, delayed-posttest design, with random

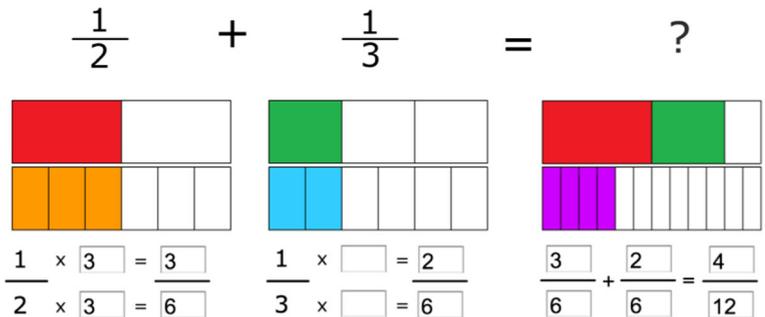


Fig. 10 Screenshot of a fraction addition tutor with an attempted implementation of grounded feedback. In each problem students are given the addends (e.g., $1/2$ and $1/3$) with red and green fraction bars representing them, and a multi-colored sum fraction showing the combined magnitudes of the addends. The student enters symbolic fractions when converting the fractions and finding the sum. The student-inputted fractions are reflected in the orange, blue, and purple fraction bars

assignment to conditions. 128 fifth-grade students completed all parts of the study. Tests consisted of four question types: fraction addition, transfer items from released standardized test questions, evaluating if a proposed fraction addition solution is correct, and items assessing pre-requisite knowledge for fraction addition. On the fraction addition items, both conditions improved, with no significant differences in improvement from pretest to posttest or from pretest to delayed posttest. On the test overall, the fraction bar condition improved more than the control condition from pretest to delayed posttest, driven by greater improvement on the transfer and evaluation items. However, the fraction bar condition had lower scores at pretest, and there was not a significant difference in scores between the two conditions at delayed posttest. Although students in the fraction bar condition improved significantly from pretest to posttest, their behavior with the tutor suggested that they could not easily use the fraction bar feedback to decide if their work was correct (the first criterion of grounded feedback). Students often indicated that they were done solving a problem even when the fraction rectangles showed that the proposed sum differed from the combined magnitudes of the two addends (Wiese 2015). This student behavior indicated a mismatch between the feedback design and the students' prior knowledge: students were not using the fraction bars as implicit verification, likely because they were not interpreting them correctly. Therefore, the fraction bar feedback was not grounded for these students. Note that the determination that the feedback was not grounded was based on process measures, not outcome measures, and that students still learned from the fraction bar tutor. The criteria for grounded feedback certainly do not include the stipulation that students must learn from the feedback for it to be considered grounded – that is an empirical question. However, there is a difference between implicit verification that students can and cannot interpret. Grounded feedback is restricted to the former.

Discussion

This paper presents *grounded feedback*, defined by four criteria: 1) Students can easily interpret the feedback to tell if it shows the right answer; 2) The feedback is intrinsic to the domain and reflects the student's inputs; 3) The feedback affords inferences on errors; and 4) The input format matches the domain learning goals. These criteria encourage designers to think about what relevant prior knowledge students are likely to have, and what accessible representations may help students link that prior knowledge to the target content. Prior work provides experimental support for grounded feedback (Nathan's ANIMATE 1998; Mathan & Koedinger's Excel tutor 2003), but controlled experiments with random assignment are limited. Further, prior work also shows that grounded feedback is not simple to implement (Rau et al. 2012; Wiese 2015). From a theoretical perspective, grounded feedback may lead to more robust learning than related forms of feedback, but sufficient empirical comparisons with sample sizes greater than 20 have not yet been conducted.

When is Grounded Feedback Applicable?

Grounded feedback can be applied to any domain where 1) students are learning to use a novel representation, 2) an intrinsic representation exists that is more accessible, and

3) responses to questions have a range of accuracy beyond right or wrong. This paper has discussed some of these domains: algebraic equations, including linear and quadratic functions (with animations, graphs, diagrams, and calculated values as intrinsic representations), Excel formulas (with the calculated value as an intrinsic representation), chemistry (with a ball-and-stick model as an intrinsic representation), fraction addition (with number lines and diagrams as intrinsic representations, which could also apply to equivalence and magnitude), and numbers (with a number line as an intrinsic representation). However, while grounded feedback could be implemented for all of these domains, we did not find controlled studies that measured learning in all of them. Other topics where grounded feedback could apply (and examples of intrinsic representations): symbolic representations for natural numbers, including counting, simple arithmetic, and rounding (number lines, chips, stacks of items); free-body diagrams in physics (graphs, freeze-frame animations of the situation, e.g., van der Meij and de Jong 2006); line graphs of functions (tables, animations for distance-time graphs); mathematical notations, such as for functions (graphs) and sigma and capital Pi (expanded algebraic expressions); equations for shapes in two and three dimensional coordinate systems (graphical visualizations).

Grounded feedback is not applicable to all topics and domains. Grounded feedback cannot be applied in domains where there is no separate intrinsic representation (e.g., geometry proofs often reference diagrams, but there is no separate intrinsic representation for the entire proof). Grounded feedback also cannot be applied to topics where the focus is conceptual change (e.g., ReLLab, Horwitz and Barowy 1994). For feedback to be grounded, students must be able to evaluate their work by comparing the current state of the feedback representation to their expectations. This requires that students have enough prior knowledge to know what the feedback should look like. When a topic requires conceptual change, students' expectations will often be incorrect, for both representations. Grounded feedback does not apply to tasks where the students' answers are limited to discrete responses which are either correct or not, with no ordering of incorrect answers where one is closer than another (e.g., identifying which arithmetic operation is required for a simple word problem).

Challenges in Designing Grounded Feedback

The main challenges in designing grounded feedback are 1) matching the feedback representation to students' prior knowledge and 2) selecting appropriate feedback features for the learning goals. As discussed above, even when the feedback representation is more concrete, students may not use it to evaluate their work (Wiese 2015). A difficulty factors assessment (c.f. Koedinger et al. 2008; Stampfer and Koedinger 2013) is a lightweight way to test if students can interpret the feedback (relative to implementing and testing the feedback in a tutor). In a difficulty factors assessment, problems of the same type are developed with different formats, or with different levels of scaffolding, and are given as a test to students in the target group. The resulting scores on each problem type indicate which are easier, and can help isolate which factors make that type of problem difficult. To test designs for grounded feedback, problems should be presented with the novel representation alone (as a baseline), with the feedback representation alone (to test that students can interpret it correctly), and with the novel representation and feedback representation together (to ensure that the

novel representation does not interfere with the interpretation of the feedback representation; Stampfer and Koedinger 2013). Further, students need metacognitive skills to evaluate and revise their work, even when interpretable resources are available. The second challenge is in selecting the domain-specific design features of the feedback. Fraction bars show a colored section representing the magnitude, and they can also show dividing lines to represent the denominator and the bounding box to represent the unit. Using or not using these features does not change how intrinsic the representation is, but will likely affect how students think about fractions. Further, in many cases, the feedback representation may not be able to reflect all responses that a student may create with the novel representation (e.g., fraction bars may not be programmed to show arbitrarily large fractions greater than 1, or negative fractions). In such cases, the designer must choose what values to show, and what to do if a student enters values outside that range (e.g., even if the system does not show fractions greater than one, it should not reinforce the misconception that those fractions do not exist). In other cases, the student may be able to create a nonsensical response with one representation that cannot be shown at all with a more concrete representation (e.g., the system of equations $x = 5$, $y = 8$, $x = y$ cannot be shown with any concrete representation because it cannot be true). The instructional designer must use other means, such as text feedback, to handle these cases.

Individual differences such as prior knowledge, metacognitive skills, and attitudes toward the domain are likely to play a role in the effectiveness of grounded feedback for individual students. Students with greater familiarity with the feedback representation or greater familiarity with the topic may attend more readily to the relevant features. The success of grounded feedback relies on students' evaluation and revision of their own work, and students with greater proficiency with these metacognitive skills may benefit more. Further, the premise of grounded feedback – that the feedback representation provides implicit verification for the novel representation – requires that the domain be internally consistent. Students who do not think the domain must be internally consistent may decide that an answer is correct for the novel representation but incorrect for the feedback representation. Finally, students who are not confident in their understanding of the domain may, upon seeing feedback for an incorrect answer, revise their conception of the underlying workings of the domain instead of revising their answers (e.g., a student doing fraction addition and seeing an incorrect answer with fraction bars could conclude that the sum of two fractions does not equal their combined magnitude, rather than concluding that the answer was wrong). However, randomized controlled studies on grounded feedback are very limited, and we have not found any that investigate individual differences.

Future Work

Future work should examine if each feature of grounded feedback is necessary, how students interact with grounded feedback, and how grounded feedback and explicit supports are best combined. In particular, the benefits of grounded feedback will likely be enhanced with a more sophisticated, AI-driven external loop (Narciss 2008). In addition to providing the grounded feedback, the system should monitor how well students are using the grounded feedback and provide appropriate meta-level support. For example, if students do not recognize that they made a mistake, the system could

direct their attention to the relevant feedback feature. Repeated failures to notice errors may indicate gaps in prior knowledge, and the system could direct the learner to remedial activities. Intelligent systems could also provide adaptive feedback based on information from multiple trials, such as suggesting that the learner try a new strategy if the student's actions have increased the magnitude of the error instead of decreasing it. Finally, adaptive systems could fade out the grounded feedback after performance reaches some threshold.

We hypothesize that students learning with grounded feedback will engage in sense making through mapping: from the input representation to the feedback representation, and from the target knowledge to their prior knowledge. In this manner, grounded feedback would provide students practice in building knowledge through inference and in checking their work using their own prior knowledge. If grounded feedback does indeed strengthen these skills, over time it may help students learn both domain knowledge and the metacognitive skills necessary to become more reflective learners.

Acknowledgements The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C100024 to WestEd, and by Carnegie Mellon University's Program in Interdisciplinary Education Research (PIER) funded by grant number R305B090023 from the US Department of Education. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198. doi:10.1016/j.learninstruc.2006.03.001.
- Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational Systems in Learning Primary Mathematics. *The Journal of the Learning Sciences, 11*(1), 25–61. doi:10.1207/S15327809JLS1101_2.
- Beckmann, S. (2004). Solving algebra and other story problems with simple diagrams: a method demonstrated in grade 4–6 texts used in Singapore. *The Mathematics Educator, 14*(1), 42–46.
- Bodemer, D., Plötzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science, 33*, 73–95.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnik & J. M. Levine (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington: American Psychological Association.
- Dugdale, S. (1992). The design of computer-based mathematics instruction. In J. H. Larkin & R. W. Chabay (Eds.), *Computer-assisted instruction and intelligent tutoring systems* (pp. 11–45). Hillsdale: Lawrence Erlbaum Associates.
- Dzikovska, M., Steinhauser, N., Farrow, E., Moore, J., & Campbell, G. (2014). BEETLE II: deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. *International Journal of Artificial Intelligence in Education, 24*(3), 284–332. doi:10.1007/s40593-014-0017-9.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–270). New York: Routledge.
- Heffernan, N. T., & Koedinger, K. R. (1998). A developmental model for algebra symbolization: the results of a difficulty factors assessment. In Gernsbacher, M. A. & Derry, S. J. (Eds.) *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 484–489). Mahwah, New Jersey: Erlbaum.
- Horwitz, P., & Barowy, B. (1994). Designing and using open-ended software to promote conceptual change. *Journal of Science Education and Technology, 3*(3), 161–185. doi:10.1007/BF01575178.
- Izsak, A. (2000). Inscribing the winch: mechanisms by which students develop knowledge structures for representing the physical world with algebra. *The Journal of the Learning Sciences, 9*(1), 37–41.

- Kalyuga, S., Chandler, P., Touvinen, J., Sweller, J., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93(3), 579–588.
- Käser, T., Busetto, A. G., Solenthaler, B., Baschera, G. M., Kohn, J., Kucian, K., et al. (2013). Modelling and optimizing mathematics learning in children. *International Journal of Artificial Intelligence in Education*, 23(1–4), 115–135. doi:10.1007/s40593-013-0003-7.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264. doi:10.1007/s10648-007-9049-0.
- Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: evidence from algebra problem solving. *Cognitive Science*, 32(2), 366–397. doi:10.1080/03640210701863933.
- Lave, J. (1988). *Cognition in practice: mind, mathematics and culture in everyday life*. Cambridge: Cambridge University Press.
- Mathan, S., & Koedinger, K. R. (2003). Recasting the feedback debate: benefits of tutoring error detection and correction skills. In H. U. Hoppe, M. F. Verdejo, & J. Kay (Eds.), *Artificial intelligence in education: shaping the future of learning through intelligent technologies, proceedings of AI-ED 2003* (pp. 13–18). Amsterdam: IOS Press.
- Mathan, S., & Koedinger, K. R. (2005). Fostering the intelligent novice: learning from errors with metacognitive tutoring. *Educational Psychologist*, 40(4), 257–265. doi:10.1207/s15326985sep4004_7.
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 745–783). Mahwah, New Jersey: Lawrence Erlbaum.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, D. M. Merrill, J. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). New York: Lawrence Erlbaum Associates.
- Nathan, M. J. (1998). Knowledge and situational feedback in a learning environment for algebra story problem solving. *Interactive Learning Environments*, 5(1), 135–159.
- National Council of Teachers of Mathematics. (2014) Equivalent Fractions [applet] Retrieved from <http://illuminations.nctm.org/Activity.aspx?id=3510> . Accessed 16 Oct 2014.
- Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, 103(2), 241–262.
- Ozgun-Koca, S. A. (2004). The effects of multiple linked representations on students' learning of linear relationships. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 26, 82–90.
- Padalkar, S., & Hegarty, M. (2012). Improving representational competence in chemistry with model-based feedback. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual meeting of the cognitive science society* (pp. 2162–2167). Austin: Cognitive Science Society.
- Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of Artificial Intelligence in Education*, 24(3), 251–283. doi:10.1007/s40593-014-0018-8.
- Powers, W. T. (1973). *Behavior: the control of perception*. New York: Hawthorne.
- Rau, M. A., & Evenstone, A. L. (2014). Multi-methods approach for domain-specific grounding: an ITS for connection making in chemistry. In S. Trausan-Matu & K. Boyer (Eds.), *Proceedings of the 12th international conference on intelligent tutoring systems* (pp. 426–435). Honolulu: Springer International Publishing Switzerland.
- Rau, M. A., Aleven, V., Rummel, N., & Rohrbach, S. (2012). Sense making alone Doesn't do it: fluency matters too! ITS support for robust learning with multiple representations. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th international conference on intelligent tutoring systems* (pp. 174–184). Berlin: Springer. doi:10.1007/978-3-642-30950-2.
- Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 41–95). Hillsdale: Erlbaum.
- Roschelle, J., Kaput, J., & Stroup, W. (2000). SimCalc: accelerating students' engagement with the mathematics of change. In M. J. Jacobson & R. B. Kozma (Eds.), *Innovations in science and mathematics education: advanced designs for technologies of learning* (pp. 47–75). Hillsdale: Erlbaum.
- Sarama, J., & Clements, D. H. (2009). "Concrete" computer manipulatives in mathematics education. *Child Development Perspectives*, 3(3), 145–150. doi:10.1111/j.1750-8606.2009.00095.x.
- Schoenfeld, A. H. (1988). When good teaching leads to bad results: the disasters of "well taught" mathematics courses. *Educational Psychologist*, 23(2), 1–22.
- Shodor. (2014) Interactivate. <http://www.shodor.org/interactivate/>. Accessed 16 Oct 2014.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. doi:10.3102/0034654307313795.

- Stampfer, E., & Koedinger, K. R. (2013). When seeing isn't believing: influences of prior conceptions and misconceptions. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1384–1389). Berlin: Cognitive Science Society.
- Suh, J., Moyer, P. S., & Heo, H.-J. (2005). Examining technology uses in the classroom: developing fraction sense using virtual manipulative concept tutorials. *Journal of Interactive Online Learning*, 3(4), 1–21.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- University of Colorado, Boulder. (2014) *PhET Interactive Simulations*. <https://phet.colorado.edu/>. Accessed 16 Oct 2014.
- Utah State University. (2014) *National Library of Virtual Manipulatives*. <http://nlvm.usu.edu/>. Accessed 16 Oct 2014.
- Uttal, D. H., Amaya, M., Maita, M. R., Hand, L. L., Cohen, C. A., O'Doherty, K., & Deloache, J. S. (2013). It works both ways: transfer difficulties between manipulatives and written subtraction solutions. *Child Development Research*, 2013. doi:10.1155/2013/216367.
- van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction*, 16(3), 199–212. doi:10.1016/j.learninstruc.2006.03.007.
- Wiese, E. S. (2015). *Toward sense making with grounded feedback*. (Doctoral dissertation, Carnegie Mellon University). Retrieved from <http://reports-archive.adm.cs.cmu.edu/anon/hcii/CMU-HCII-15-104.pdf>.
- Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education*, 33(2–3), 153–169. doi:10.1016/S0360-1315(99)00030-5.
- Yerushalmy, M. (1991). Effects of computerized feedback on performing and debugging algebraic transformations. *Journal of Educational Computing Research*, 7(3), 309–330.