Why Direct Instruction Earns a C- in Transfer

Dan Schwartz, Stanford Univ.
http://AAALab.stanford.edu/

Most teachers have experienced the frustration of students doing well on chapter tests but forgetting in other contexts. Many times, students did not forget; they were confused. They could not figure out which of their concepts applied to the new situations. Whitehead coined the expression "inert knowledge" to describe peoples' failure to apply their learning when they should. In behavioral research, it is known as the transfer problem. The transfer problem raises three questions: Why does it happen? Should we care? What can we do?

Why Direct Instruction Fails to Transfer

Successful transfer depends on learning at least two things: the relevant concepts or skills, and the situations to which they apply (2). In one study, advanced psychology students demonstrated solid memory of clinical theories, but once they met patients, they could not determine which theories applied (3). There are many good instructional techniques for improving memory, but teaching the situations that call for those memories is a different matter.

When teaching about situations, students need to learn the deep structures that exist across many instances (4). Experts recognize deep structures. Physicists, for example, categorize both spring and inclined plane problems as cases of energy conservation, whereas novices see them as problems about different devices (5).

A prevailing cognitive explanation for failed transfer is that students cannot get past the surface or incidental features of instructional instances. For example, if students learn about density by studying gold, they may not spontaneously transfer the concept of density to water.

Research has found that students are more likely to transfer if instructional examples are abstract and relatively free of surface details (6, 7, 8). For instance, students who learned to compound quantities using abstract math transferred the formula to physics, but students who learned the same formula through applied physics did not transfer to math (9).

Despite this evidence, the surface details of instructional examples are not the primary cause of failed transfer. Instead, the clash between surface features and transfer is an inadvertent outcome of direct instruction. Typical instructional methods for science and math give students the formula first, and then have students complete practice problems. The strength of such "tell-and-practice" methods is that students quickly memorize the formula. The pitfall is that students focus on the formula and never attend to the generalizable structure of the problem situations (10). They are only left with memories for the formula and the obvious surface features of the problems.

A study with 8^{th}-grade students demonstrates the negative effects of tell-and-practice for learning deep structure (11). The lessons focused on density and speed. A shared deep feature is their ratio structure (D=m/V, S = d/t). In the first lesson, students in the Tell-and-Practice treatment were directly taught about density and the formula. Afterwards, they received three paired cases. For each pair, they had to apply the density formula. Figure 1a shows that each pair represented a company that had two busses with the same clown density. In the Invent treatment, the students were not told about density. Instead, they were asked to invent a quantitative index for the crowdedness of the clowns.

[Figure 1 about here – Clown Task and Memory Results]

A day later, students redrew the cases from memory. Students in both treatments remembered similar numbers of surface features. However, the Tell-and-Practice students did not reproduce the ratio structure as often as the Invent students (Figs 1b-c). For the Tell-and-Practice students, direct instruction deflected attention from the deep structure, so only easy-to-notice surface features were encoded. For the Invent students, the surface features did not interfere with encoding the deep structure.

These differences affected subsequent transfer. Over the next few days, students completed three more pictorial activities on density and speed, maintaining their instructional treatments. The Tell-and-Practice students further received a lecture on the importance of ratio for density, speed, and other topics in physics. Afterwards, all students took a transfer test that depicted springs pressing against plates of different sizes. Their task was to describe the surface pressure. This was a new topic for the students, but it also involves a ratio structure (springs over area). Tell-and-Practice students used a ratio to describ the surface pressure at 41% the rate of the Invent students (Fig 2a). This limited transfer is notable given that these students had completed four tasks that analogously used ratios to describe physical phenomena (12); and, they had been explicitly told that ratios were the common structure and are important more generally in describing physical situations.

[Figure 2 about here – Transfer Results]

The poor transfer was not due to direct instruction *per se*, but rather because the direct instruction pre-empted processing the deep structure. After completing the first transfer test, the Invent students also received direct instruction on the formulas and ratio, and both treatments practiced on word problems. Two weeks later, all students completed a delayed transfer test on the spring constant – determining the stiffness of trampoline fabrics. The Tell-and-Practice transfer rate was 36% of the Invent rate (Fig. 2b). Direct instruction did not interfere with transfer for the Invent students.

Direct instruction is important, because it delivers the explanations and efficient solutions invented by experts. To gain this benefit without undermining transfer, direct instruction can happen after students have engaged the deep structure, per the Invent condition. Figure 2c shows that the Invent students performed just as well on a subsequent test of word problems about density and speed. Direct instruction becomes problematic when it shortcuts the appreciation of deep structure. Across conditions, students who encoded the deep structure of the clown problems were twice as likely to transfer. It is just that fewer students in the Tell-and-Practice condition encoded the deep structure, because they had received direct instruction too soon.

A survey of transfer studies published in the past five years revealed that 75% used tell-and-practice for both control and treatment conditions (13). This impressive number indicates how deeply direct instruction is entrenched in both teaching and research. With only direct instruction, one would predict the average finding that students do not transfer because they only encode surface features. Nevertheless, despite being the norm, direct instruction earns less than a C grade. Instruction can do better. As the current research indicates, the problem of surface features for transfer is not a baked-in human frailty. Rather, poor transfer is driven by an instructional context that replaces the need to think about deep structure with a premature solution to memorize.

Why We Should Care

Hatano distinguished two courses of expertise: routine and adaptive (14). Routine expertise thrives in highly stable situations, for example, an assembly line where the environment conspires to remind people what to think and do. The "routine" of the task creates little need for transfer, and the premium is on replicating behaviors with greater speed and precision.

Adaptive expertise is important for novel and dynamic situations, where people might adapt to variability rather than perseverate on familiar routines (15, 16). Because students consistently face new ideas and situations within school and beyond, teaching towards adaptive expertise is valuable. Transfer is relevant to adaptive expertise, because people need to remember old ideas in new situations, and then adapt them to learn.

Appropriate instruction can prepare students for future learning and adaptation (17). For example, most 8^{th}-graders already have qualitative intuitions about surface pressure and spring stiffness, but the Invent students developed an extra edge. Their appreciation of ratio structure helped them to begin learning that these intuitive concepts are relational. In contrast, the Tell-and-Practice students focused on single factors. So, rather than improving their understanding of spring stiffness, Tell-and-Practice students thought about how far a spring stretches and disregarded the forces involved.

The Invent students demonstrated transfer that was useful for learning, but necessarily constrained by the research design. The practice and transfer cases were intentionally similar to highlight the limitations of tell-and-practice for encoding and transfer. A multi-week study on teaching statistics better demonstrates the potential of preparing students for future learning by enabling transfer (18). High-school students learned statistics through Tell-and-Practice or Invent treatments. The end of their final exam included a novel transfer problem on normalizing data and that depended on procedures they had not learned. Half of the students in each condition received a worked example in the middle of their test that showed the relevant procedures. These students had to follow the worked example to solve a problem on the same page, which they all did quite well. However, Figure 3 shows that only the Invent students subsequently used the procedure to solve the novel transfer problem. The Invent students exhibited a double transfer. They transferred learning from the invention activities to the worked example, and then transferred from the worked example to solve the novel problem on normalizing data.

[Figure 3 about here – Inventing Prepares Students to Learn]

The statistics study indicates one reason for educators to care about transfer. Appropriate instruction prepares students to transfer so they can learn more deeply from the expository presentations that run through much of education. For example, college students learn more from lectures and readings when they first work with relevant data compared to when they write a summary of a chapter that explains the same data (19).

Of course, not all education concerns issues of transfer. Some skills are only used in stable contexts that provide strong cues to their application. Word decoding, for example, is always applied to texts that are spaced into words that read from left-to-right. Decoding lessons should emphasize high efficiency, so that later on, students can allocate cognitive resources to more variable aspects of the task such as reading for meaning. For some skills, routine expertise is the instructional goal.

In math and science, instruction cannot exhaust all possible situations. Transfer and adaptation are important. Although automaticity is important for some facts such as "2 x 3 = 6," real situations rarely come with formulas attached, so students need to learn to recognize the relevant deep structures. Moreover, the cumulative curricula of math and science mean that students should build a base of knowledge on deep structures from which future learning can grow and adapt.

What to Do about Transfer

There are many alternatives to tell-and-practice including problem-, project-, and inquiry-based instruction (20, 21). The inventing activities are different because they strategically precede standard pedagogy rather than replace it. They can be brief additions, because they use highly structured contrasting cases to support rapid induction. The contrasting cases, like glasses of wine or crowded busses side-by-side, help people notice not only what differs, but also what constitutes the common deep structure (22, 23). Student experimentation can also generate contrasting cases, but only if students can design the contrasts well (24).

Contrasting cases are only part of the story, however, because both the Tell-and-Practice and Invent students worked with them. The nature of the task is also important. During invention tasks, students author a compact model, graph or formula to distill the cases. For example, given a variety of molecules, students might model a cell membrane that permits only certain molecules to pass. Working towards a single way to handle all the cases leads students towards the deep structure that generalizes across them. The task of creating a compact representation also prepares students to appreciate the symbolic work done by the representation in the expert solution.

Students may initially feel uncomfortable not knowing the "sure way" to get a high grade. However, this passes once students engage the task and iterate through inventions that progressively handle more of the cases and deep structure. Students do not have to invent the canonical solution to reap the benefits of inventing. Either way, they will be better prepared to understand the solution when it is explained to them (19). Furthermore, student inventions make illuminating topics for class discussion. (For more guidelines, see (25)).

Instructional models aside, educators can use transfer assessments as a way to identify areas where school is not providing students what they need for future learning. While the 8th-graders in both conditions performed the same on standard word problems, the transfer assessments revealed how tell-and-practice can preempt the search for structure and undercut progressive learning.

In the current milieu of high-stakes testing, standardized assessments largely measure routine expertise; namely, efficient recapitulation. If educators want students to become adaptive, innovative citizens who keep learning through changing times, current assessments do not fit. A better fit would map students' trajectory towards adaptive expertise. Ideally, assessments would examine students' ability to transfer, particularly for new learning. Such assessments would include resources for learning <u>during</u> the test (for example, a simulation that students can freely manipulate). Education needs fresh measures that index more than routine expertise, so that instruction can contribute to the goal of creating an adaptive, 21st-century workforce. (26)

References

(1) Whitehead, A. N. (1929). The aims of education. New York: MacMillan.

(2) Bransford, J. D., Franks, J. J., Vye, N. J. & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 470-497). New York, NY: Cambridge University Press.

(3) Michael, A. L., Klee, T., Bransford, J. D., & Warren, S. (1993). The transition from theory to therapy: Test of two instructional methods. Applied Cognitive Psychology, 7, 139- 154.

(4) Greeno, J.G., Smith, D.R. & Moore, J.L. (1993). Transfer of situated learning. In: D. K. Detterman and R. J. Sternberg (Eds.), Transfer on trial: Intelligence, cognition, and instruction (pp. 99-167). Norwood, NJ: Ablex.

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

(6) Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (April, 2008). The advantage of abstract examples in learning math. Science, 320, no. 5875, pp. 454-455.

(7) Uttal, D. H., Liu, L. L., & DeLoache, J. S. (1999). Taking a hard look at concreteness: Do concrete objects help young children learn symbolic relations? In C. S. Tamis-LeMonda (Ed.), Child psychology: A handbook of contemporary issues (pp. 177–192). Philadelphia, PA: Psychology Press.

(8) Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. Journal of the Learning Sciences, 14, 69-110.

(9) Bassok, M., & K. J. Holyoak (1989). Interdomain Transfer Between Isomorphic Topics in Algebra and Physics. Journal of Experimental Psychology, 15, 153-166.

(10) Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. Memory & Cognition, 19, 543–557.

(11) See Supporting On-Line materials for methods and data analyses.

(12) Gick, M. L., & Holyoak, K. J. (1983). Schema Induction and Analogical Transfer. Cognitive Psychology, 15, 1-38.

(13) See Supporting On-Line materials for methods and results.

(14) Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, and K. Hakuta (Eds.), Child development and education in Japan (pp. 262-272). NY: Freeman.

(15) Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. Journal of experimental psychology: Learning, memory, and cognition, 14, 510-520.

(16) Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. Cognitive Science, 22, 319-346.

(17) Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), Review of Research in Education , 24, 61-101. Washington DC: American Educational Research Association.

(18) Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. Cognition & Instruction, 22, 129-184.

(19) Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. Cognition & Instruction, 16, 475-522.

(20) Barron, B. J., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., Bransford, J.

D., & CTGV. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. Journal of the Learning Sciences, 7, 271-312.

(21) Linn, M. C., Lee, H-S., Tinker, R., Husic, F., & Chiu, J. L. (August, 2006). Teaching and assessing knowledge integration in science. Science, Vol. 313. No. 5790, 1049-1050.

(22) Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment. Psychological Review, 62, 32-51.

(23) Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. Child Development, 67, 2797–2822.

(24) Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. Psychological Science, 15, 661-667.

(25) More guidelines and examples for inventing activities may be found at http://www.cwsei.ubc.ca/resources/files/Teaching_Expert_Thinking.pdf.)

(26) This material is based upon work supported by the National Science Foundation under grants SLC-0354453 & REC-0231946. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank Kimi Schmidt for help in executing the empirical study; Dylan Arena for his help in the survey of the transfer literature; and, Lindsay Oishi for her editorial talents.

Figure Captions

Figure 1. How Different Forms of Instruction Affect Student Encoding of Surface and Deep Features.

Figure 2. Tell-and-Practice Yields Low Transfer and No Special Benefit for Basic Computation.

Figure 3. Appropriate Instruction Prepares Students to Learn and Transfer (Study 1 from (18)). After completing Invent or Tell-and-Practice activities, students received a final exam with a difficult transfer problem near the end. Half of the students in each condition received a worked example embedded in the test. Both groups solved the worked example problem over 85% of the time. However, only the Invent students learned from the worked example so they could better solve the transfer problem.

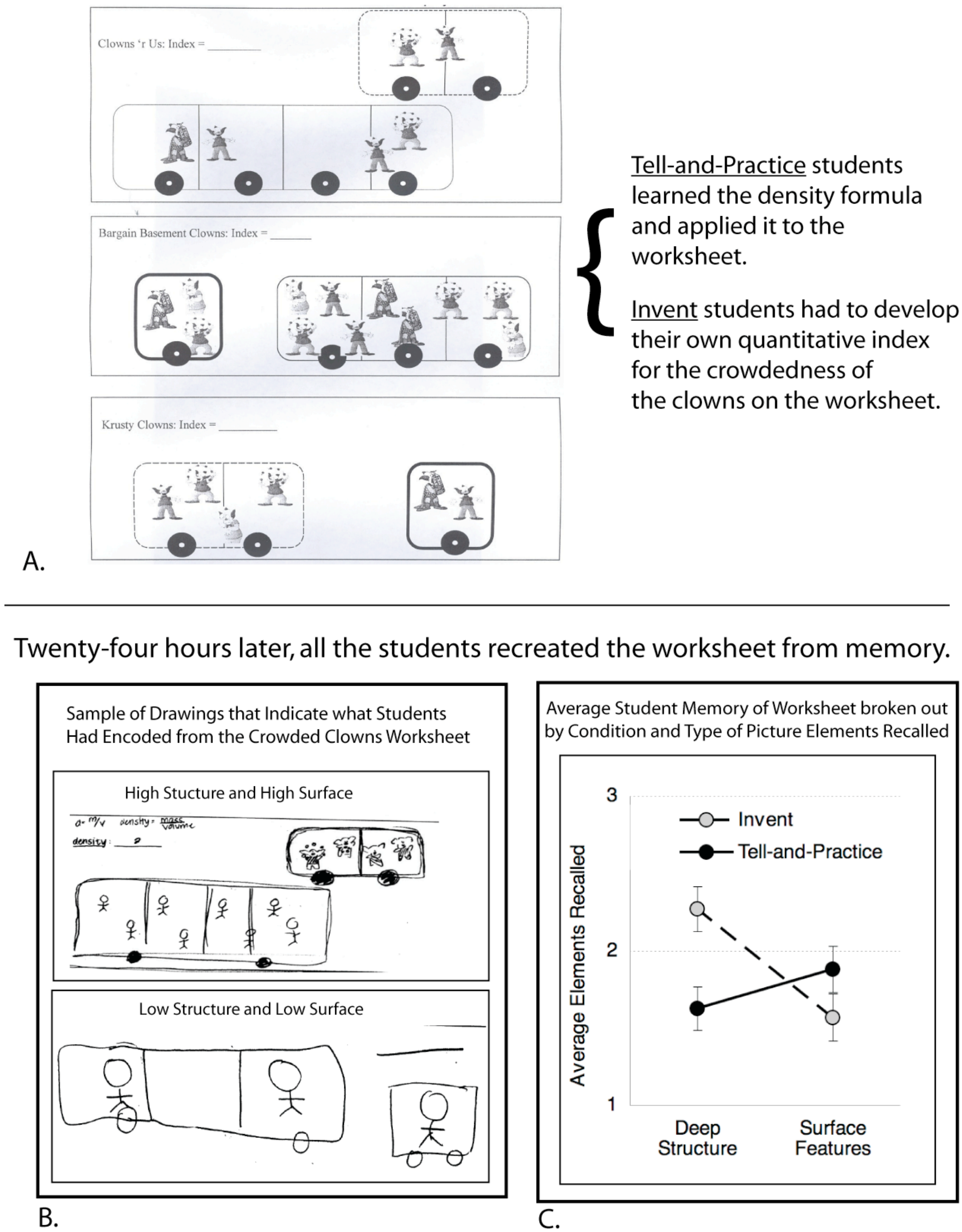


Figure 1. How Different Forms of Instruction Affect Student Encoding of Surface and Deep Features.

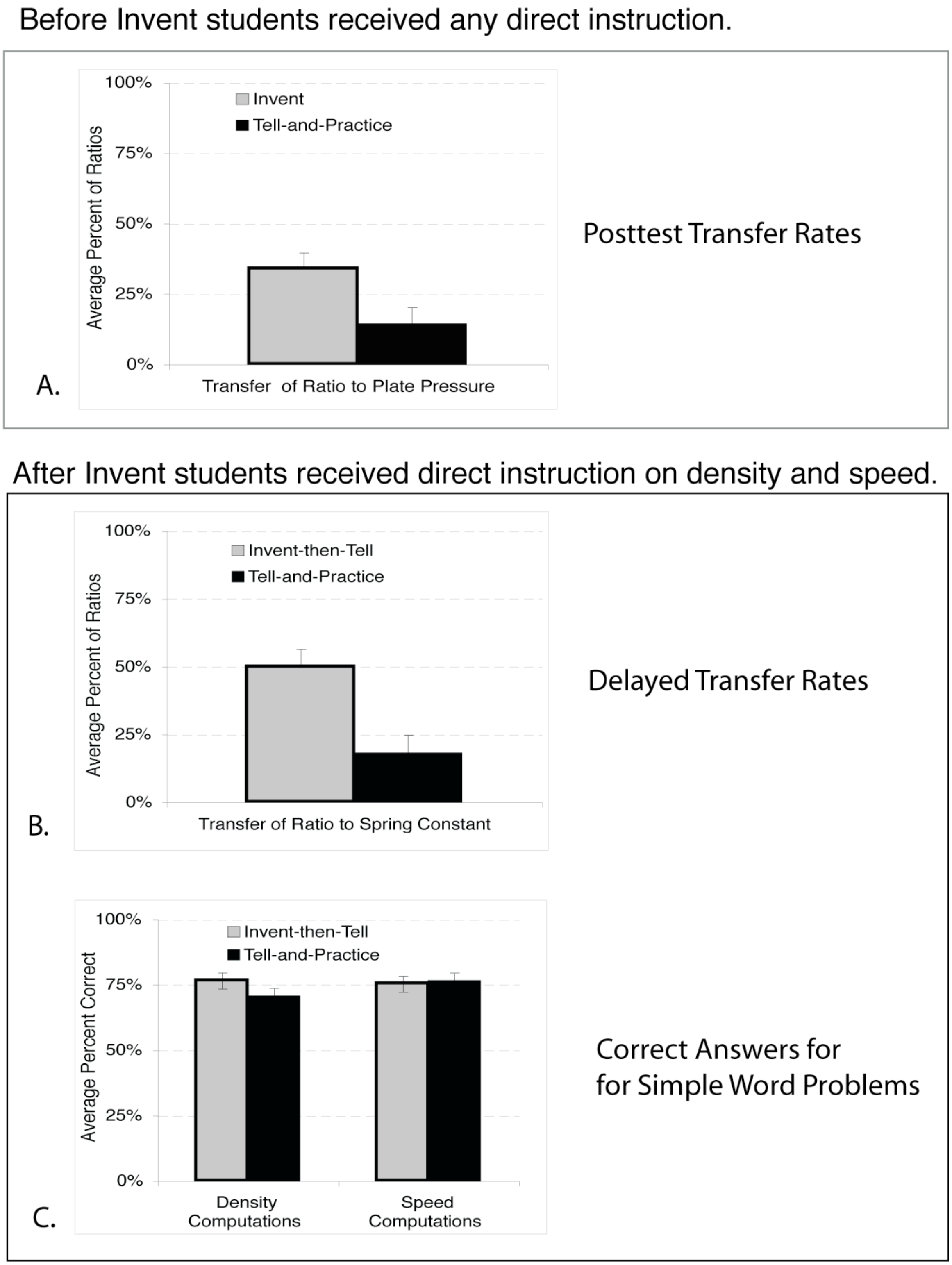


Figure 2. Tell-and-Practice Yields Low Transfer and No Special Benefit for Computation.

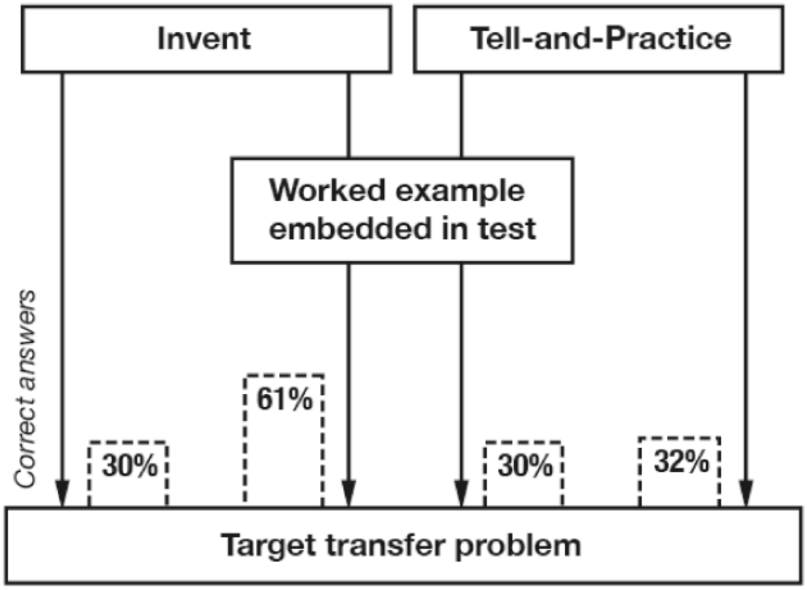


Figure 3. Appropriate Instruction Prepares Students to Learn and Transfer (Study 1 from (18)). After completing Invent or Tell-and-Practice activities, students received a final exam with a difficult transfer problem near the end. Half of the students in each condition received a worked example embedded in the test. Both groups solved the worked example problem over 85% of the time. However, only the Invent students learned from the worked example so they could better solve the transfer problem.

On-Line Appendix: Why Direct Instruction Earns a C- in Transfer

The first section describes the empirical study on learning physics. The second section describes the survey of the transfer literature.

SECTION 1:

STUDY ON EFFECTS OF DIRECT INSTRUCTION FOR ENCODING AND TRANSFER

This study examined the effects of direct instruction on encoding and transfer. Direct instruction took the form of explanation plus a worked example, and then a set of cases for practice. It is labeled Tell-and-Practice. As a control condition, students completed a form of instruction where they had to invent a quantitative representation to characterize the same sets of cases. It is labeled Invent.

The research was intended to demonstrate the peril of direct instruction when it shortcuts processing of deep structure. Students in the Tell condition did not notice the deep features, and they showed poor transfer. The design of the research was not intended to support strong claims about the benefits of the Invent condition for transfer. The students in the Invent condition had to look at data to devise a compact representation that worked for all the cases during the learning phase. This was the same task required at transfer, though this was not stated explicitly in the transfer instructions and the physics content was different. Learning to find structure in data is of high significance in science, so the transfer tasks have face validity. However, if a transfer task had taken the form of analyzing data from their own experiments or some home activity, it is conceivable the students may not have transferred the use of a ratio. Other research, discussed in the main text (17, 18), has demonstrated the value of invent treatments for subsequent transfer across more varied contexts.

[Figure A-1 about here: Study Timeline]

Figure A-1 outlines the timeline of the study, with key instructional and assessment components. Both treatment groups completed the same four worksheets on speed and density. Both conditions received a lecture on the mathematical structure underlying all the topics, namely, ratio. However, the Tell-and-Practice students heard the lecture near the beginning of the study, whereas the Invent students heard the lecture after completing all the worksheets. For the Invent students, the final lecture also taught them about the density and speed concepts and their relevant formulas, which the Tell-and-Practice students learned prior to each activity. Afterwards, students in both conditions practiced the formulas on simple word problems. Assessments targeting different hypotheses were spread across the study. They are described in the Methods.

The lessons were intended to help students learn about the mathematical structure underlying many phenomena in 8^{th}-grade physics. They were not intended to be comprehensive treatments of density or speed. For example, there were no lessons on the implications of different densities for physical behaviors like buoyancy. Ratio was selected as a target structure, because it is a key invariant across many physical descriptions. Children of this age are typically able to reason about ratio and proportion (equivalent ratios), but they still tend to rely on a single dimension of magnitude. The focus on magnitudes of only one factor can lead to confusion, for example, between density and weight, or between acceleration and speed. Ideally, by exposing students to the power of this simple mathematical structure, it will support future transfer to new physics concepts. Moreover, given a longer experience, it may cultivate the attitude that mathematics can be an ally rather than an enemy when it comes to science.

Methods

Participants

8th-grade students (n=128) in four different science classes at the same school participated near the end of their academic year. The students attended a highly diverse middle school with over six ethnicities and average-to-low standardized test scores. The study used stratified random sampling to assign conditions. Within each class, students were ranked by course grade, and then assigned alternately to one treatment or the other. The students for each treatment moved into separate rooms to receive instruction. Thus, the experiment was implemented four times across the classes. To ensure fair balance in instruction, four instructors (three researchers and the classroom teacher) were rotated across classes and treatments. For example, on one day, instructor X might work with the Tell-and-Practice treatment for period 2 and the Invent treatment for period 3. On another day, instructor X might do the Invent treatment for period 2 and the Tell-and-Practice treatment for period 3.

Design

During 4 days of instruction, students learned about density and speed through either the Tell-and-Practice or Invent methods. After one day of lessons, a free recall task assessed student encoding of the initial cases. Students in the Tell-and-Practice treatment then heard a lecture on the significance of ratio. After two more days of worksheets, students received the first transfer test on surface pressure. Afterwards, students in the Invent treatment received direct instruction on density, speed, and the importance of ratio. All students were then given the opportunity to practice on word problems, followed by a quantitative test on density and speed. Two weeks later, students completed a delayed test on quantitative and qualitative reasoning, plus a second transfer item on the spring constant.

Materials and Procedures

Due to other school activities, weekends, and a national holiday, the study was not conducted on consecutive days. Moreover, some days had shortened class periods due to school assemblies and so forth. These did not have a differential impact on the implementation of either treatments or classes.

[Figure A-2 about here. Instructions for clown task]

Day 1 (Mon). Students were assigned to their respective conditions and moved into separate rooms by condition. The specific rooms varied from period to period due to space constraints at the school. Students in the Tell-and-Practice group received an explanatory sheet that included a worked example on computing density as the number of objects in a space. This is Figure A-2a. The teacher had the class read the worksheet aloud and discussed any questions. The Invent group received a different explanatory sheet shown in Figure A-2b. It explained the idea of an "index" such as miles per gallon. It also explained the task of inventing an index of clown crowdedness and why one might do such a thing (crowded clowns are grumpy clowns). The sheet did not provide any examples for how to compute an index.

The clown worksheet included three companies, each with two busses of clowns (Figure 1 in the main text and in part in Figure A-1). There was a space for inserting the index (for

Invent) or the density computation (for Tell-and-Practice) for the pair of busses for each company. The clown cases were created for two reasons. The first is that the absurd situation was unlikely to cue the Invent students to use the density formula if they had learned it elsewhere; if they had applied the density formula from memory, this would have made the Invent treatment much the same as the Tell-and-Practice treatment. The second reason for the clown cases is that they readily permitted building in deep and surface features. There were six surface features built into the cases that were useful for detecting surface memory. These are described in the coding section below. There were three deep features, namely, the proportional ratios for each of the 3 companies. These were useful for detecting whether students had encoded the ratio structure of the cases.

Students were allowed to work individually or in groups of up to three students. This format was consistent with the teacher's regular practice. The elective grouping was implemented for all classroom exercises. In general, students worked in pairs, and they often switched partners from day to day. For the assessments, students worked individually under standard testing conditions.

<u>Day 2 (Tues).</u> The start of the class began with a surprise, free recall test. Students in both treatments received an identical sheet. The sheet stated, "*Yesterday, you received a sheet that had an activity with clowns and busses. Use the space to redraw how the sheet looked the best you can. Work alone please."* Students had as much time as they wanted to redraw the original sheet (typically in the 10 minute range).

After completing the memory test, students in the Tell-and-Practice students received a lecture that explained how ratios are an important structure in physics. They saw examples of relevant equations (D=m/V, S=d/t, a=F/m) on the white board. There was discussion about each. Students in the Invent condition did not receive an initial lecture on ratio in science (this was reserved until after completing all the exercises).

Students in both groups then received the popcorn popper assignment (partially shown in Figure A-1). This represented speed as events per time. The different task goals and explanatory sheets remained the same for both groups across these types of lessons for the remainder of the study: The Tell-and-Practice group continued to receive an explanatory sheet for each lesson describing the formula and a worked example, and the Invent group received a cover story explaining the need for an index (minus the definition of the index concept used in the first lesson). Also, as before, both classes read the sheets aloud and there was sufficient discussion to clarify the task (and formula for the Tell-and-Practice condition). Students then received the identical popcorn popper work sheet, where their task was to compute the speed or invent an index that indicated the rate of popping for different machines. As before, there was a space for reporting the speed/index for each company.

Day 5 (Fri). Students in the Tell-and-Practice condition received a sheet that explained density as a measure of continuous mass over volume along with a worked example. Students in the Invent condition received a sheet that explained that gold comes in different degrees of purity and that gold of lower purity has been mixed with lighter metals. Their task was to find an index of gold purity for the cases, while the task for the Tell-and-Practice condition was to compute the density using the given formula. Cases appeared as a set of cubes of various volumes and masses (see Figure A-1). After completing the cases, the students received a new worksheet that contained cases involving the speeds of different cars. Students worked on speed as a function of distance over time, instead of events over time as before. This task was slightly different from before, because the students in both conditions had to decide which cars came from the same company. (In the prior activities, cases from the same company were juxtaposed.)

[Figure A-3 about here. Transfer tests.]

Day 8 (Mon). Students in both groups received the first transfer test. Working individually, they received a sheet that showed four rectangular "aerosol" cans organized into pairs (Figure A-3a). The instructions explained that aerosol cans use pressure to push out the liquid. To allow pictorial representation and to simplify the problem, this pressure was portrayed as a set of springs that pushed on plates at the bottom of each can. Their task was to "describe" the plate pressure that each company uses in their aerosol cans. The word "describe" had not been used previously in either treatment. The instructions did not include the words "index" or "compute".

Afterwards, the Invent condition finally received direct instruction on the appropriate formulae, conventional names (i.e., density and mass), and the significance of ratios. The lecture captured in one session, what the Tell-and-Practice students had heard spread over several sessions. Similar to the Tell-and-Practice lecture on ratio, the Invent lecture indicated that ratios are an important form of mathematics for understanding physics, and that many phenomena have a ratio structure. The instruction then explained the density and speed formulas and how they applied to discrete and continuous cases, which the Tell-and-Practice students had learned before each activity. Afterwards, they received a practice worksheet of simple word problems and straight computation. The worksheet included the formulas at the top. Representative problems include: (a) *Time = 3 hours. If speed = 36 mph, what is the distance traveled?* (b) *Brenda packs*

120 marshmallows into 4 soda cans. Sandra packs 300 marshmallows into eleven soda cans. Whose soda cans are more densely packed?

The Tell-and-Practice condition had received the summary lecture prior to working on the cases, as well as the specific sheets that explained density and speed. Therefore, they had approximately 10-15 minutes of extra time, while the Invent condition finally learned the conventional solutions. This extra time was filled by providing the Tell-and-Practice students with the same word problems as the Invent students, plus a few extras. Overall, the Tell-and-Practice treatment provided very minimal savings over the Invent treatment in instructional time.

<u>Day 11 (Thur).</u> Students in both classes received a standard quantitative and one-step word problem test. It had problems of the same form as the computation worksheet from the day before, though there were no equations at the top of the sheet and students had to remember and apply the appropriate formulas. There were three problems for each of density and speed.

<u>Day 29 (Mon).</u> Students received a delayed test that covered the topics in this study as well as other material in biology that they had learned in the intervening days. The test included an abbreviated set of quantitative and qualitative problems, and a transfer problem. The qualitative problems asked questions such as, "*If the weight of the cube were less, would the density be higher or lower?*"

The transfer problem showed four different trampolines, different numbers of people on each trampoline, and different distances of stretch (Figure A-3b). There was no mention of companies or organization into explicit pairs to ensure that these surface features were not responsible for condition differences at transfer. The sheet explained: "*Trampolines are made with mats using different fabrics. Stiffer mats make the trampoline bouncier. Determine the* *stiffness of the mat fabric for each trampoline."* The stiffness of the fabric was an application of the spring constant.

Coding

Free Recall of the Clowns worksheet. Students attempted to redraw the clown worksheet from the day before. Drawings were coded for the inclusion of deep features and surface features. For deep features, they received 1pt for each pair of busses that had the same ratio (and were not redundant with other busses). The ratios did not have to be the same as the original worksheet, but they had to be proportional for pairs of busses. The maximum possible score was 3 pts. For surface features, they received one point for the inclusion of each of the following: elaborating the clown features to look like the original sheet; drawing the clowns on the lines separating the compartments of a bus; drawing different style lines for at least two busses; including wheels that did not correspond in a principled way to the number of compartments; including the name of one of the companies; and, recreating incidental text features such as the spot for writing in their names. The maximum score was 6. Surface and deep features were chosen to be orthogonal. For instance, the spatial location of the busses could not be used as a feature in the coding. Students who omitted buses would receive a lower structural score, but they would also necessarily receive a lower surface score. If allowed, this would have violated the independence of the two measures.

For this and all measures, two independent reviewers coded 25% of the data, equally sampled from both conditions. One coder took responsibility for scoring the remaining 75% of the data. For the free recall data, the coders had a single 1 pt disagreement on deep features, and two 1 pt disagreements on surface features.

Transfer problems. For both the pressure and spring transfer items, the key question was whether students would use ratios to describe the phenomena. For each problem, students received 1 point for each correct ratio (inversions of the numerator and denominator were considered acceptable answers). Given four cases for each transfer problem, a maximum score is 4. There were no coder disagreements for either transfer measure.

A secondary level of coding looked for the following in the transfer problems: (1) Did students correctly rank the cases in terms of pressure or stiffness? Many students in the Tell-and-Practice condition explicitly ranked the cases, even though they did not compute ratios. (2) How many different types of elements of the problem situation did they quantify? (3) Did students incorporate 0, 1, or 2+ factors in their description? Qualitative statements such as there are "more springs for each cube" would count as two factors, whereas "more springs" would count as one factor.

Across these three secondary measures there were a total of 6 coder disagreements cumulative across the two transfer items (280 coding events).

Quantitative and qualitative problems. The first quantitative test comprised multiple problems. Each problem was given a pair of 1 point codes. For the computation code, students received 1 point, if they either gave the correct answer or set up the correct equation (calculation errors were not counted against the students). For the units code, 1 point was awarded if students included the correct units in their answers. For the units, students received credit even if they used the wrong exponent for spatial measures (e.g., mm^2 instead of mm^3).

The delayed test contained quantitative and qualitative word problems. For the qualitative problems, students received 1 pt for giving the correct qualitative answer (e.g., "increases"). To avoid over-weighting the quantitative scores when aggregated with the qualitative scores, the students had to get both computation and units correct to receive a single point for the quantitative problems. There were 408 quantitative items coded for reliability for each test. There were fewer than 2% disagreements for both tests. For the qualitative problems on the delayed test, there was no coder disagreement.

Results

Memory Effects

Students in the Tell-and-Practice (Tell) condition did not recall the deep structure of the cases nearly as well as the Invent (Invent) students; M=1.57 and M=2.26, respectively. There were no significant differences in their memory for surface features; Tell M = 1.88 and Invent M = 1.63. These results indicate that surface features do not necessarily block encoding of deep structure. The equivalent memory for surface features also demonstrates that there was not some incidental implementation difference that might have caused the Tell students to have a generalized poorer memory.

On the day of the memory test, 122 students with informed consent were in-class. A repeated measures analysis included the two types of feature memory (surface v. deep) as a within-subject factor and treatment as a between-subjects factor. There was no main effect of treatment, $F(1,120)=2.4$, MSe=1.24, $p > .1$. There was also no main effect of feature type on memory, though it should be recalled that the memory scores are not on equivalent scales, $F(1,120) = 1.2$, MSe=1.39, $p>.25$. However, there was a strong treatment by memory type interaction, $F(1,120)=9.9$, $p < .005$. The interaction is driven by the differences in the deep features. Taken separately, the treatment shows a large effect for deep features, $F(1, 121)=11.0$,

$p < .005$, but no effect for surface features, $F(1, 121)=1.5$, $p > .2$. This entails that surface and deep features were not trading off. There were no significant correlations between surface and deep features, such that across all participants the correlation was r = -.08. These effects were consistent across all four classes.

To get a sense of the number of students who remembered the structure, structure scores were converted into a binary, structure-present versus structure-absent coding. Using the criterion of a perfect structure score, 59% of the Invent students versus 34% of the Tell students had encoded the structure. A more lenient coding considers whether the student recalled two or more of the three ratio pairs. By this analysis 80% of the Invent students encoded the structure compared to 53% of the Tell students. Either way, more students in the Invent condition paid attention to the structure as indexed by their reconstructive memory.

Transfer Effects

To demonstrate transfer, students had to include ratios in their characterization of the transfer cases. The Invent condition showed superior transfer compared to the Tell condition for both the immediate transfer test and the delayed transfer test (see Figure 2 in the main text). Due to an implementation error, data from one class for the first transfer test could not be used. The analyses including the first transfer test exclude the lost class, but all classes are included in the delayed transfer analyses.

A repeated measures analysis used time of test as a within-subject factor (immediate v. delayed) and treatment as a between-subjects factor. There was a very strong treatment effect, $F(1, 93)=13.4$, MSe=3.83, $p < .001$. There was a marginal effect of test, $F(1, 93)=3.9$, MSe=2.11, $p < .06$, which probably occurred because the delayed transfer problem was easier (i.e., less candidate features for consideration in the description). Despite the descriptive evidence that the Invent treatment did relatively better on the delayed transfer test, there was not a treatment by time-of-test interaction, $F(1, 93)=0.9$, MSe = 2.11, $p > .3$. It is possible that transfer improved descriptively for the Invent students, because it occurred after their "telling" and the practice word problems. The delayed transfer test was also analyzed separately, so it was possible to include all four classes. The effect of treatment on the delayed test remained very strong, $F(1,125)=13.6$, MSe=3.45, $p < .001$. The advantage for the Invent condition was consistent across all classes for both transfer tests.

Additional evidence more closely binds the noticing of structure in the clowns cases with subsequent transfer performance. Although there were intervening lessons between the clown task and the delayed transfer tasks, it is possible to see if the structural recall had any relation to transfer. Students' clown memory was coded as successful, if for over half of the cases, they correctly included proportionate ratios. Students' transfer was coded as successful if they included ratio descriptions for over half of the cases on the transfer test. This analysis was confined to the delayed transfer test, because it permitted inclusion of the lost class and the scores were not nearing floor levels. When students failed to encode the ratio structure of the clowns, they only had a 23% chance of solving the delayed transfer item. When students had encoded the ratio structure of the clowns, they had a 46% chance of solving the delayed transfer item, $\chi^2(1, N=121)=6.0$, $p < .05$. There was no interaction with condition, indicating that students who found structure in the Tell condition were also more likely to transfer. The difference is that the Tell students were less likely to encode the structure from the clown cases, and presumably all the subsequent density and speed cases.

One potential concern with the focus on ratios is that students may have been learning math, but they may not have been learning anything useful about physics. The ratio-focused transfer tests may not be a useful indicator of students' physical understanding of surface pressure and spring constants. One way to address this concern is to show that students who did not use ratios were unable to qualitatively rank the cases in terms of the target physical properties. For example, for the delayed trampoline transfer problem, did the students correctly rank the trampolines in terms of the stiffness of the fabric? Without the use of ratios, students had a very difficult time understanding the physical property qualitatively. For the first transfer test, students who used ratios for at least half of the cases correctly ranked the relative surface pressure on the plate 91% of the time. Students who did not use ratios succeeded 5% of the time, χ^2 (1, N=96)=66.1, $p < .001$. Similarly, for the delayed transfer item, students who used ratios for over half of the cases correctly ranked the relative stiffness of the fabric 96% of the time. Students who did not use ratios only succeeded 1% of the time, χ^2 (1, N=126)=113.7, $p < .001$. Students who did not use ratios tended to rank the cases based on the value of a single feature, for example, the number of people on the trampoline. Thus, without ratios, students had very little chance of forming a precise qualitative understanding of the physics, because they focused only on one part of the physical situation. As an analogy, students often replace density with weight, because they do not appreciate that density is relational.

Table A-1 provides more general summary statistics broken out by condition. "Physical Conclusion" refers to whether students correctly ranked the cases according to the target physical property (e.g., stiffness of the trampoline mat). "Quantities Extracted into Numbers" refers to how many of the features of the problem did they quantify (e.g., number of people on the trampoline mat). "Factors Included in Description" refers to how many key dimensions of the problem students included in their descriptions quantitatively or qualitatively (e.g., stretch distance of trampoline would be one factor, and number of people on the trampolines would be a second factor).

Table A-1. Student conclusions and handling of problem features in case descriptions.

	Surface Pressure (Aerosol)		Spring Constant (Trampoline)	
	Invent (n=49)	Tell (n=47)	Invent (n=65)	Tell (n=61)
Physical Conclusion (% of students)	35%	15%	51%	25%
Quantities Extracted into Numbers (Avg.)	1.3	0.8	1.2	0.8
Factors Included in Description (Avg.)	1.2	1.0	1.4	1.1

Note: Based on a MANOVA analysis, all treatment comparisons of dependent measures within for each transfer test are significant at $p < .05$ or better, excepting the comparison of "Factors Included in Description" for the Surface Pressure transfer test.

Word Problems and Computation

The immediate quantitative posttest included three questions involving density and three problems involving speed. Some problems required computing speed or density, whereas other questions gave density or speed, and other parameters like mass or time had to be computed. Each item was coded separately for the accuracy of the equation setup and for the inclusion of correct units. Figure A-4 shows that there were no reliable differences between conditions.

[Figure A-4 about here. Quantitative and Qualitative problem solving results.]

A repeated measures analysis crossed within-subject measures of physics concept (density or speed) by answer element (computation or units code), which was crossed with the between-subjects factor of treatment. There was no main effect of treatment, $F(1, 116)=0.1$, MSe=0.21, $p > .8$. There were no interactions with treatment either, all p's > 0.1. *A priori*, it seemed possible that the Tell condition would do better at labeling the units, because they had more instruction on which parameters went into the denominator and numerator, and they practiced with the cases. In contrast, the students who invented solutions to the cases could flip the numerator and denominator without affecting the quality of their index for the tasks at hand. There was a descriptive advantage for the Tell treatment for units, but it did not approach significance. Evidently, the "telling" in the Invent condition plus a few minutes of practice was sufficient to match the Tell condition, which had been explicitly told the formulas before each case.

More generally, students made more errors on units than computations, $F(1, 116) = 149.4$, MSe = 0.07, $p < .001$. Students tended to leave out the units, which is a fairly common problem in science instruction. There was also a physics concept by computation/units interaction, $F(1,116)= 35.0$, MSe= 0.04, $p < .001$. Students were more inclined to leave out units for the speed problems. Conceivably, because distance, time, and speed are more intuitive for the students, they projected the units onto the numbers, and did not bother to label them. This, of course, is speculative.

The delayed test did not reveal any differences between conditions either. There were three density and two speed questions. For each concept, the first question was a quantitative word problem, and follow-up questions asked what would happen to one parameter if there was qualitative change in another parameter (e.g., if weight increases, what happens to density). When analyzed separately in a MANOVA, there were no treatment effects for any of the questions. To present the data aggregated, each student received an average percentage correct for density and for speed by combining quantitative and qualitative scores. The two percentages were repeated measures with treatment as a between-subjects variable. There was no effect of treatment as a main effect, $F(1,124)=1.7$, MSe=0.1, $p > .15$, nor in interaction with physics concept, $F(1,124)=1.4$, MSe=0.05, $p > 0.2$. There was a strong effect of physics concept, F(1,124)=94.8, MSe=0.05, p<0.001. Students had more trouble with the density problems. One explanation is that the speed word problem began, "A truck travels 100 miles in 2 hours…" This maps into the division order in S=d/t, and the numbers are easy, which explains the improvement over the posttest. The density word problem began, "A 4 cm^3 cube of metal weighs 12 grams…" This reverses the order of the division in D=m/V, which led some students to divide grams into cubes.

Discussion

The direct instruction did not lead students to encode the structural features that are important for transfer. This was demonstrated by the poorer memory of structural features on the free recall of the clown worksheet. It was also demonstrated by the poor performance on the transfer items. In contrast, the inventing task did lead students to encode the deep structure more effectively as indicated by the memory test. And, they transferred better.

With respect to the poor transfer for the Tell-and-Practice students, it is important to note the high similarity and cues of the cases used at learning and transfer. They had a distinctive pictorial quality; they were arranged on the page in a common format; they were all clearly from the same "researchers"; and, they all included readily quantifiable elements that comprise a ratio, which the students were told about and had to use during the learning phase. The materials in the learning phase included discrete and continuous versions of both density and speed creating four analogs, and analogical cases are a strong way to promote transfer. Yet, the Tell-and-Practice students infrequently induced the ratio structure during learning or at transfer. Students in both conditions who noticed the ratio structure of the clown cases transferred significantly more often. The problem for the Tell-and-Practice students was that the direct instruction led them away from the deep structure of ratio, presumably, because they were paying more attention to following the formula and taking each case one at a time.

The Invent and Tell-and-Practice students performed the same on the qualitative and quantitative word problems. This indicates that the students had equivalent knowledge of the procedures. The poor transfer of the Tell-and-Practice condition cannot be attributed to a general lack of knowledge. These students had learned the ratios, what they had not learned was the properties of situations that call for ratios.

SECTION 2:

SURVEY OF TRANSFER ARTICLES PUBLISHED BETWEEN 2003-2008

The goal of this literature review was to determine the percentage of transfer studies that used direct instruction for all treatments. The goal was not to determine what conditions have shown positive and negative evidence of transfer. In fact, the point was to get some estimate of how many transfer results need to be reconsidered given the current data.

The preceding transfer results showed that transfer has a great deal to do with the broader context of instruction. In many transfer studies, direct instruction has been taken for granted while researchers manipulate other variables like motivation, goal orientation, quality of visual and auditory presentation, abstractness of materials, and so forth. Thus, researchers have made claims about the variables that effect transfer and the psychological processes behind it. Given the current findings, these claims may have to be circumscribed to conditions that involve direct instruction. The practical implication is that a good deal of instruction that has been predicated on the results of transfer studies may have led to the design of sub-optimal curriculum. Of course, this cannot be demonstrated based on a literature survey. Rather, the goal is to get some estimate of what percentage of the transfer literature may require reconsideration.

A three filter process produced a good-sized sample of the educationally relevant literature that has used transfer as an outcome measure. The first filter was to search the ISI Web of Knowledge for studies published in the last 5 years (2003-2008) using the search term "transfer." This produced a variety of articles from many disciplines, for example, the transfer of chemicals across cell membranes. To produce articles relevant to human behavior, the next filter included only those articles indexed by the terms "thinking," "procedure," "learning," or "psychology." This yielded 350 articles. The third filter was done by hand. To be included in the final set, the articles had to report original research involving normal human populations learning science or math topics with transfer as an outcome measure. For example, learning the control of variables strategy is relevant to science learning whereas learning the Tower of Hanoi problem is not. Learning about the principles of radar is relevant to science learning, whereas training at using a specific radar system is not. Six people reviewed separate subsets of the articles. If an individual found an ambiguous case, it was directed to the other five reviewers who decided by committee and blind to instructional method.

All told, the sampling strategy yielded a total of 70 articles comprising 136 unique authors. Other search strategies and criteria would yield somewhat different samples, but the large sample size ensures that the results are likely to generalize to other search strategies. (The resulting list of articles is available upon request.)

Each study in each article was coded as a direct-instruction-only study if all of the treatments in the study used direct instruction. Direct instruction included forms of instruction where students were told or shown the key methods or ideas prior to engaging in their work. For example, in one study, the authors compared the effects of a human-like voice versus a computer-sounding voice on a measure of transfer. In both treatments, the instruction involved following direct instruction and worked examples, after which students practiced. The direct-instruction-only appeared in research that used relatively long stretches of direct instruction; research that decomposed tasks into steps with repeated cycles of short direct instruction then practice; and, research that used worked examples that students then followed. There were other variants as well.

Across the articles, 75.0% of the individual studies were direct-instruction-only. This high value indicates how greatly direct instruction suffuses current instruction and research. In fact, the method of instruction is often taken for granted. Looking at the article abstracts, only 58.6% described the instructional treatments in enough detail to determine if they used direct instruction or some other pedagogy. Instead, many of the abstracts focused on the variables being manipulated, and ignored the instructional context in which those manipulations took place.

Some articles included more studies than others. It is possible to correct for the potential over-weighting of some research publications that included very many studies. A separate computation determined the percentage of studies within an article that were direct-instruction-only. The results are about the same. The average number of studies per article is 1.64 (SD =

1.0), and the average number of direct-instruction-only studies per article is 1.23 (SD = 1.2). The average percent of direct-instruction-only studies per article is 72.1%. Whatever results these studies are producing, it is important to recognize that the results may not generalize to other forms of instruction.

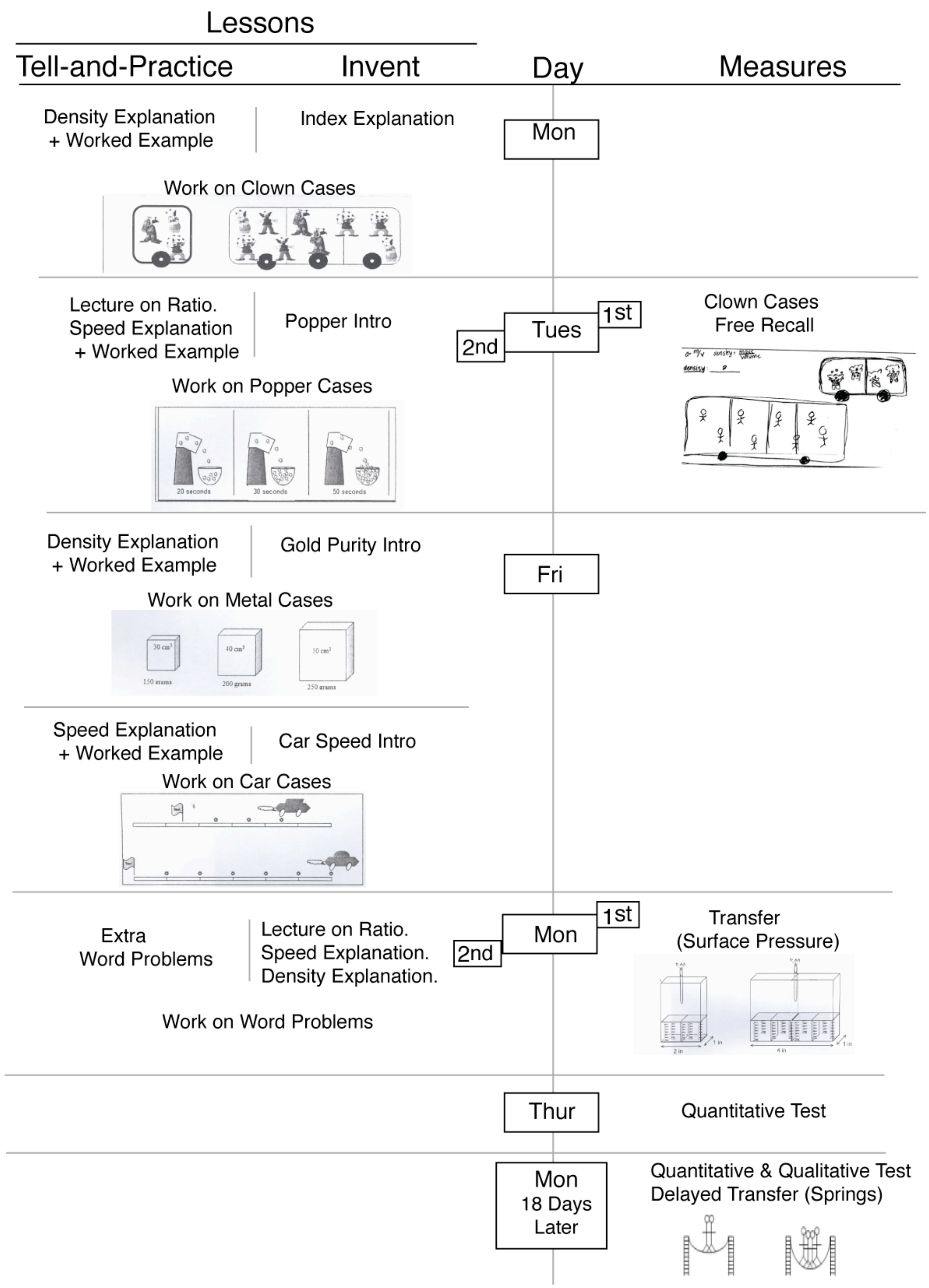


Figure A-1. Timeline of Major Study Activities.

A.

FINDING **DENSITY**

Density is how much stuff is packed into a space. Density can be the number of people in a room, the density of feathers in a pillow, and many other things.

Density is very important in chemistry. Density is a property of ***matter***. Gold is denser than carbon, because more matter is packed into each atom of gold compared to each atom of carbon.

When working with density, the trick is to use the simple equation:

$$D = \frac{M}{V} \quad \text{or} \quad \text{Density} = \frac{\text{Mass}}{\text{Volume}}$$

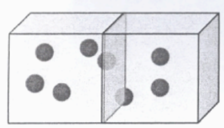
Density is a measure of the mass of a substance per unit of volume.

Sometimes ***Mass*** can be found by counting.

Volume is the amount of space. Volume is harder to find, because a volume can take many shapes – a sphere, a balloon, a bottle.

To make it easier, we will tell you the volume. We will measure it in cubes.

In the example below, there are two cubes. There are 8 objects spread across the cubes. Density is the average number of objects per unit of volume.



Density = mass / volume
= 8 objects / 2 cubes
= 4 objects / cube
= 4 objects per cube

On the next page, compute the density for each bus of clowns.

B.

Inventing an Index

An **index** is a number that helps people compare things.

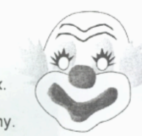
Miles per gallon is an index of how well a car uses gas.
Batting average is an index of how well a baseball player hits.
Grades are an index of how well you are doing in school.
Star rating is an index of how efficient an electrical appliance is.

We want you to invent a procedure for computing one kind of index.

THE CROWDED CLOWNS INDEX

Companies send clowns to parties, circuses, amusement parks, sporting events, and so on.

To get the clowns to the event, each company packs the clowns into a bus.
Some companies make the clowns more crowded than other companies.



The more crowded the clowns are, the grumpier they will be.

People who order clowns want to know a company's crowded clown index.

Invent a procedure for computing a crowded clown index for each company.

RULES FOR THE INDEX

1. The same company always crowds the clowns the same amount, no matter how many clowns get ordered. So a company only gets a single crowded clown index.

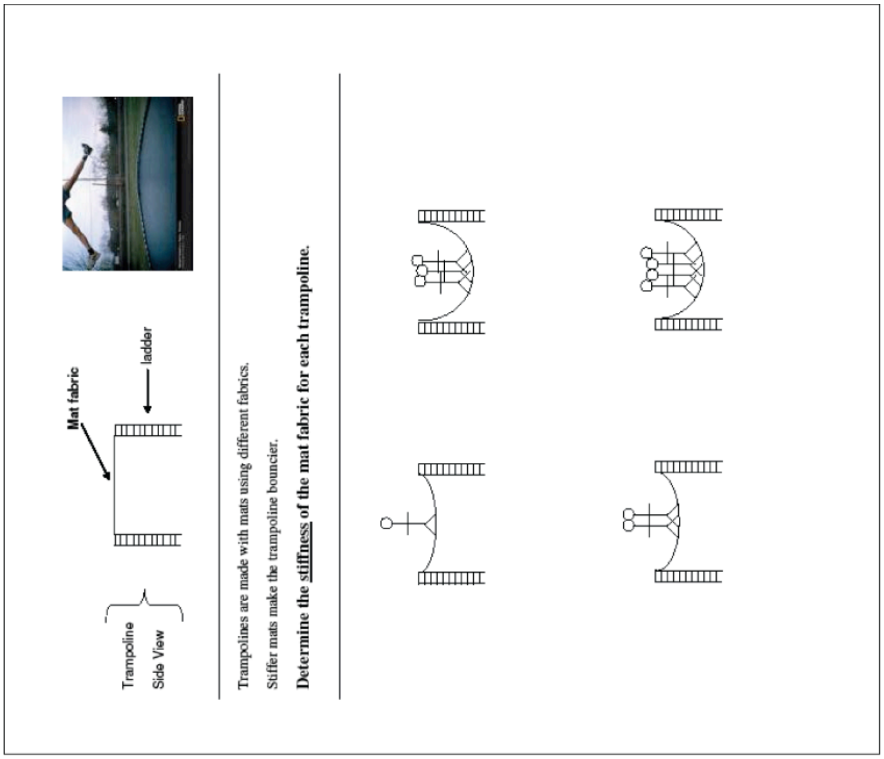
2. You have to use the exact same procedure for each company to find its index.

3. A big index value should mean that the clowns are more crowded. A small index number should mean that the clowns are less crowded.

Good luck!

Figure A-2. Instructions Prior to the Clown Cases.

B.



A.

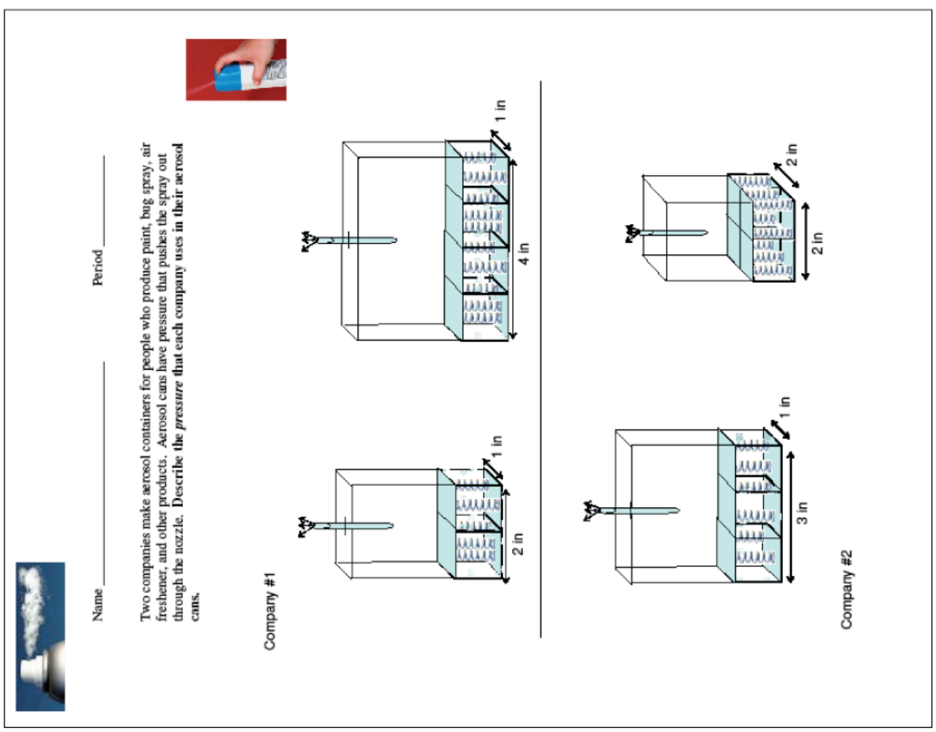
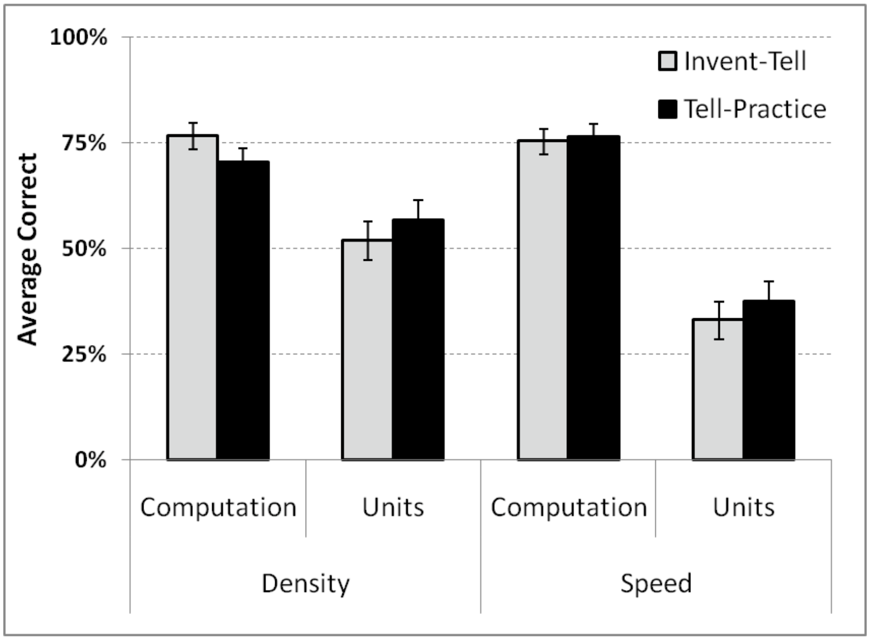
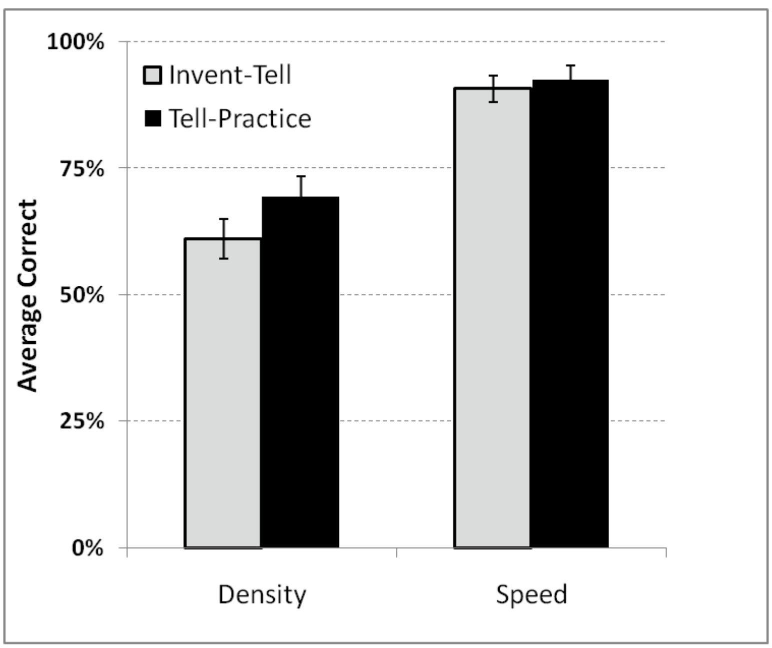


Figure A-3. Transfer Items for (A) Immediate, and (B) Delayed



A. Immediate.



B. Delayed.

Figure A-4. Student Performance on One-Step Problems on Immediate and Delayed Test.