

Assessment of Vertical Transfer in Problem Solving: Mapping the Problem Design Space

Joshua Von Korff, Dehui Hu and N. Sanjay Rebello

Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506-2601

Abstract. In schema-based theories of cognition, vertical transfer occurs when a learner constructs a new schema to solve a transfer task or chooses between several possible schemas. Vertical transfer is interesting to study, but difficult to measure. Did the student solve the problem using the desired schema or by an alternative method? Perhaps the problem cued the student to use certain resources without knowing why? In this paper, we consider some of the threats to validity in problem design. We provide a theoretical framework to explain the challenges faced in designing vertical transfer problems, and we contrast these challenges with horizontal transfer problem design. We have developed this framework from a set of problems that we tested on introductory mechanics students, and we illustrate the framework using one of the problems.

Keywords: transfer of learning, problem solving, problem design, vertical transfer, construct validity

PACS: 01.40.Fk

INTRODUCTION

Researchers and teachers often design problems to assess their students' knowledge and ability to transfer this knowledge to new contexts. These problems may be invented on a whim, developed based on pedagogical or research experience, or tested and refined using a cyclical process. Any problem can be well-designed or poorly-designed for the purpose of assessing a given type of knowledge or ability. In this paper, we present a theory of problem design.

We will distinguish between vertical transfer problems (VTPs) and horizontal transfer problems (HTPs), which measure students' knowledge or abilities in different ways. Our goals are to:

- 1) show that VTPs are more difficult to design well and HTPs are easier to design; and to
- 2) propose a classification of "threats" that interfere with the design of physics problems.

Vertical and horizontal transfer are defined by Rebello et al. [1]. According to the definition, vertical transfer means that students "construct a schema or mental model on the spot" or choose the most productive model from a number of options. Horizontal transfer situations, on the other hand, "[do] not require students to construct or to even choose between competing schemas or mental models." Rather, horizontal transfer merely requires the activation of a robust schema or model in a new situation.

As Rebello [1] notes, the same problem can involve vertical transfer or horizontal transfer for two different students. However, in this paper we will focus on the mindset of the instructor or researcher who is making use of the problem. For instance, if an instructor intends an exam question to test whether students can apply a procedure that s/he imagines they "should" already have learned, we consider this to be a horizontal transfer problem (HTP). If, on the other hand, the instructor intends to test the student's ability to invent new procedures, then it's a vertical transfer problem (VTP).

This definition suggests that HTPs usually test practical *procedures*, which an instructor thinks of as learning goals of the instruction, or which may be useful components of more complex procedures. VTPs often do not. There is an old joke that goes:

Alice: I wish I had enough money to buy an elephant.

Bob: Why do you want an elephant?

Alice: I don't. I just wish I had that much money.

The elephant is something that is not desired for its own sake, it is just a sign or symbol that Alice has enough money. In VTP design, we want the student to have enough "money" (understanding) to buy an "elephant" (solving the VTP).

Our focus on the instructor's frame of mind (rather than the student's) is helpful in this context, because we wish to consider the *construct validity* of the problem as an assessment. Construct validity is "the

evaluation of the extent to which a measure assesses the construct it is deemed to measure.” (Strauss and Smith, [3]) Construct validity of an assessment is relative to the intent of the educator or researcher who makes use of it.

Messick [4] describes three *threats* to construct validity: construct underrepresentation, construct-irrelevant easiness, and construct-irrelevant difficulty. In this paper, we propose five types of construct validity threats (CVTs), based on a subdivision of Messick’s categories. This categorization has been informed by our semester-long study of students’ difficulties with differentiation and integration in a physics context, described in the Methodology section.

CONSTRUCT VALIDITY THREATS

We will first define our five categories for validity and illustrate those using examples in the context of a simple thought experiment – a fictitious problem design task. We imagine that we are creating a HTP to assess students’ understanding of two-dimensional motion under the influence of gravity. A projectile is to be shot into the air with a known angle and initial velocity, and the students should tell us its range. To do this, they should first find the time of flight by solving a quadratic equation for the vertical motion; then they should plug that value of flight time into the equation for the horizontal motion. We will describe the five CVTs related to this HTP scenario and suggest why these threats may be easily avoided or ignored when designing HTP. In the Results section, we will use the same CVT categories to explain why problem design of a VTP is much more challenging.

CVT1: *Underrepresentation*: the problem does not measure the full extent of the skill that is to be tested. For instance, we can give one problem in which the projectile is fired upward and another in which it is fired downward into a cave. It may be that no single problem can test both of these sub-skills. On the other hand, at least this list of (two) sub-skills is finite, manageable, and not too messy. In the case of VTP, we will see that the skills to be assessed are more difficult to precisely define or describe.

CVT2.a: *Construct-irrelevant difficulty with respect to solution*: Sometimes the student does not know how to solve the problem, and the reason has nothing to do with the conceptual knowledge we aim to assess. For instance, if we named the initial velocity “ v_0 ” when prior problems in the course had named it “ v_i ,” then the student might be confused by this change in notation. This confusion represents a student’s lack of notational knowledge, but it is not relevant to the assessment goals. However, it is

always possible to design a “fair” HTP, relying only on the procedures and notation that have been taught in the course. VTPs are more likely to require deeper understanding and thus are more susceptible to CVT2.a.

CVT2.b: *Construct-irrelevant difficulty with respect to implementation*: For instance, if the student doesn’t have a calculator, then finding the sine of 37° is hard. The student knows what to do, but cannot do it. Of course, any real instructor would allow the calculator, so this example is contrived. In general, any HTP is practical (not an “elephant” problem), so the student’s implementation of the solution should not be a fundamental barrier to problem design.

CVT3.a: *Construct-irrelevant easiness with respect to cueing*: the student may access the solution too easily, as a result of a “cue” in the problem statement. For instance, if the problem statement talks about “range” and “trajectory,” these technical terms could cue the students to use certain equations, whereas some students might be helpless without the cues.

But HTPs are only concerned with assessing a student’s ability to apply a schema. Therefore, many designers of HTPs would not be too concerned about these cues, especially if it seemed inconvenient to write the problem statement without the word “trajectory” or “range.”

CVT3.b: *Construct-irrelevant easiness with respect to alternative method*: The students may have memorized the range equation, $R = (v_0^2 \sin 2\theta)/g$, and might use it instead of the procedure we want them to use. The alternative method is correct, but we might decide it does not match our assessment goal.

To counter this threat, we should assign a problem in which the projectile gets fired from a point that is at a different vertical level than the final point, such as firing it off a cliff. Usually, if a procedure is *practical*, then at least some problem exists for which the procedure is *optimal*.

In summary, it seems as though most of the CVTs for this HTP are not serious concerns. Most can be fixed easily. We think this is true of many or most HTPs for the reasons mentioned above. In particular:

- 1) HTPs rely on techniques, notations, and conventions that are woven into the fabric of the traditional physics curriculum. With a VTP, students are wandering in unfamiliar territory (CVT2.a).
- 2) The HTP designer is less concerned with whether students will *access* the right schema too easily and more concerned with whether the student can *apply* that schema (CVT3.a).
- 3) HTPs assess the student’s use of *practical* techniques, which should be both serviceable (CVT2.b) and optimal (CVT3.b).

METHODOLOGY

We completed a series of controlled trials with participants taken from an introductory mechanics course for engineers at Kansas State University. Of 262 students attending the course, 140 chose to participate in the research. In this paper, we will discuss a subgroup of 47 students who worked on a lesson about differentiation in a kinematics context. We will discuss the CVTs that arose when the students attempted a post-test after this lesson.

Forty-seven students attempted this post-test. These students were split into four sessions that met on different days. Students' time was divided into: a written pre-test (5 minutes), a lesson (50 minutes), and a post-test (30 minutes). The lessons included a seven or eight-page-long written worksheet for students to solve in groups of two to four students each. Students were given worked solutions when they completed each part of their worksheets. The pre- and post-tests were identical for the most part, but the latter were administered with both paper and online components. The online component made use of Mastering Physics, which could provide the student with a series of hints. In all cases, students were instructed to write down any thoughts that came to mind and were told that they would be graded more on the clarity and depth of the ideas expressed in their writing rather than on the correctness of their answers.

The post-test asked students to estimate the acceleration at a particular time, given a position versus time graph, $x(t)$. We anticipated that students would first find the instantaneous velocities i.e. slopes of the $x(t)$ graph at two nearby times and then use these velocities to find the acceleration using the formula $a = (v_2 - v_1)/(t_2 - t_1)$. This is an "elephant problem" that is meant to find out how well the student understands how to use a graphical representation of position versus time to find velocity and acceleration. Further, this problem also assesses whether students understand the concept of approximation. One ("treatment") version of the lesson gave students targeted practice on these issues, whereas the other ("comparison") version used kinematics problems from a popular physics textbook.

RESULTS

Having explored the CVT for a thought experiment HTP in a previous section, we will now discuss the CVT for our transfer task, a VTP. We contend that quantitative VTPs are challenging to design because they face serious construct validity threats.

CVT1: Underrepresentation: This CVT asks the questions: What did we want to assess? Did we assess that construct from all perspectives, or did we just examine one feature out of many?

With this VTP, we intended to assess students' understanding of the relationship between the graph and equation representation and their understanding of approximation as it relates average to instantaneous velocity and acceleration. We hoped that students could use this knowledge to construct our preferred solution to the transfer task.

Our transfer task certainly fell short of measuring these targeted constructs in their entirety. The "understanding" targeted by our VTP is a broader and fuzzier idea than the "ability to apply a specific schema" measured by an HTP. For instance, we could test the student's "understanding of the relationship between graph and equation" from many other perspectives. We could show the students a graph of some motion (say, a damped oscillator) and ask them to guess what equation $x(t)$ could have led to this graph. Students who succeed at the acceleration task would not necessarily succeed at the oscillator task or vice-versa. These problems test different aspects of "understanding representations." It would be hard to argue that any one problem measures all aspects of a student's understanding.

CVT2.a: Construct-irrelevant difficulty with respect to solution: This CVT asks: Apart from the construct to be measured, did the students possess all necessary knowledge and make all necessary assumptions?

One substantial issue arose for this CVT. A few students attempted to plug the average velocity, rather than the instantaneous velocity (slope of the tangent line) into the equation for the average acceleration, $\Delta v/\Delta t$. In doing so, they risked using the wrong time interval. The students' procedure requires four times: t_{1a} , t_{1b} , t_{2a} , and t_{2b} . Of these, t_{1a} and t_{1b} are needed to find the first average velocity v_1 , and t_{2a} and t_{2b} to find v_2 . But it is incorrect to find the average acceleration using $(v_2 - v_1) / (t_{2b} - t_{1a})$, especially when $t_{1b} = t_{2a}$. This is a subtle point, and our 50 minute lesson did not prepare students to understand it.

CVT2.b: Construct-irrelevant difficulty with respect to implementation: This CVT asks: If students understood the correct solution method, were they able to implement it? The transfer task, while solvable in principle, requires some degree of dexterity and precision with a ruler to find the slope and with counting fractions of square edges on graph paper. Some students selected times that were a bit too far apart, leading to an average acceleration that did not closely approximate the instantaneous acceleration.

CVT3.a: Construct-irrelevant easiness with respect to cueing: This CVT asks: Did the students base their

solution method on construct-irrelevant hints or cues in the problem statement? A student might make use of cues that we do not want students to rely on all the time, even though these cues were useful in this particular context.

The biggest concern is that by placing the transfer task immediately after a lesson which was designed to help the students find slopes and make approximations on graphs, we were strongly cueing students to use these techniques to find the acceleration. We accepted this price as a feature of the format we chose for our experiment.

In addition, our computerized hint system provided minor and major hints, one at a time. The first minor hint informed students that the answer would involve an approximation. The major hint showed students a picture of the two tangent lines they would need. These hints permitted students to continue working on problems even after they became stuck. But working with the data from these students may present a CVT, especially for students requiring major hints.

CVT3.b: *Construct-irrelevant easiness with respect to alternative method*: Many alternative methods existed to solve our problem. For example some students assumed (correctly) that the depicted function $x(t)$ was a cubic, and therefore they could sketch a linear acceleration $a(t)$. When in a later session we used a Gaussian $x(t)$, the success rate declined dramatically, since this alternative method was no longer applicable.

We were able to detect many of these threats by asking students to show their work, but this kind of remedy requires time, effort, and subjective analysis on the part of the grader. More importantly, we will never know how the students “would have” solved the problems in the absence of the threats, especially in the absence of CVTs 3.a,b.

Table 1 below shows the number of student responses that were affected by each CVT. Note that an individual student may be associated with multiple CVT, or no CVT. Further, the problem and lesson varied; so this table combines multiple scenarios.

TABLE 1. CVT detected from our transfer task. The bottom row is *not* the total of previous rows; see text.

CVT	# of students affected
CVT2.a	3
CVT2.b	8
CVT3.a (major hint)	2
CVT3.b	10
Correct, no CVT3 (or minor hints only)	7
Total number of students	N = 47

CONCLUSIONS & FUTURE WORK

We have proposed a classification of construct validity threats affecting problem design of quantitative physics problems. We expect that this classification will be useful to researchers and educators who wish to consider the conditions that must be met in successful problem design.

Our framework explains why quantitative questions may be inadequate to define a valid vertical transfer problem. A qualitative component in a problem can especially help with CVT2.a and CVT2.b. By showing their work, students are able to demonstrate their knowledge in spite of some construct-irrelevant difficulties. More research could be completed to determine the relevance of these construct validity threats to other techniques, such as clinical interviews, teaching interviews, and multiple choice tests with distracters.

We have based this framework on Messick’s categories [4] and our experience with three vertical transfer problems, one of which is presented in this paper. It is likely that more experience could lead to further classification of the CVTs. More work is also needed to further improve the framework such that it characterizes a wider variety of problems.

ACKNOWLEDGMENTS

This material is based in part upon work funded by the National Science Foundation under grant 0816207. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect views of the National Science Foundation.

REFERENCES

1. N. S. Rebello et al, “Transfer of Learning in Problem Solving in the Context of Mathematics and Physics,” in *Learning to Solve Complex Scientific Problems*, edited by D. Jonassen, Lawrence Erlbaum Associates, New York, 2007, p. 223-246.
2. D. K. Detterman, “The Case for the Prosecution: Transfer as an Epiphenomenon” in *Transfer on Trial: Intelligence, Cognition, and Instruction*, edited by D. K. Detterman and R. J. Sternberg, Ablex Publishing Corporation, New Jersey, 1993, p. 5
3. M. E. Strauss and G. T. Smith, *Annual Review of Clinical Psychology* **5**, 1-25, 2009.
4. S. Messick, *American Psychologist* **50(9)**, 741-749, 1995.