DEVELOPMENT AND ANALYSIS OF A BASIC PROOF SKILLS TEST

Sandra Merchant[a] and Andrew Rechnitzer[b]

[a] Carl Wieman Science Education Initiative, Department of Mathematics, University of British Columbia
[b] Department of Mathematics, University of British Columbia

Abstract: *We have developed a short (16 question) basic skills test for use in our institution's transition-to-proof course that assesses basic skills required to succeed in such a course. Using this test in our core introductory proof course, we have found that students are generally deficient in a number of skills assumed by instructors. In addition, using this test as a pre/post-test we have found that in this course students are learning some concepts well, but that learning gains on other concepts are much below desired levels. Finally, administration of the test to students in a higher level course has allowed us to assess retention of these skills. At this preliminary stage these skills appear to be retained into higher-level proof courses, but more data collection is needed, as well as a more extensive instrument to assess proof skills, rather than simply basic logic and comprehension.*

Keywords: Classroom research, Transition to proof, Retention, Basic skills

Introduction:

Our institution has a typical transition-to-proof course (Math 220) intended to bridge students from the computationally intensive calculus stream of courses to upper division proof-intensive courses. As is generally the case in such a course, students encounter a great deal of difficulty and instructors express a large degree of frustration with the poor learning outcomes. One source of this frustration appears to be a mismatch between students' incoming skills and those assumed by instructors. We have therefore endeavoured to identify these skills and assess them in the incoming population. An instrument we have developed to assess these skills is the "basic proof skills test" (BPST). This is a short (16 question) multiple choice test developed over several terms through a combination of instructor interviews, student interviews, and iterative development of test items, following the procedure outlined by Adams & Wieman (2011). The test is still under development, but at this stage it consists of 4 questions on algebra and graphing (hereafter the "Precalculus" content) and 12 questions on logic, mathematical quantifiers, and mathematical definitions (hereafter the "proof skills" content).

We developed this test to be used as a pre/post test in the core introductory proof course at our university with two main goals: (1) to assess students' skills at the start of the course, and hence to tailor instruction to address deficiencies and build on key skills and (2) to track learning gains for these skills over the term of the course. In addition, administration to students in a higher level course can allow us to assess retention of these skills. Over the past year we have administered the test to several sessions of the course, as well as in our introductory analysis course. Here we present the results of these and describe their implications on our teaching in these courses as well as on further improvement of the test for the future.

Methods:

We began with interviews with past instructors of Math 220, as well as with instructors of downstream courses (proof-intensive 300-level courses for which Math 220 was a pre-requisite). These were largely unstructured interviews where the discussion focused on (a) identifying what proof skills students were expected to have at the start of each of these courses, and (b) what common difficulties instructors have observed. Upon completion of these interviews, we constructed an initial version of the test that focused on the most basic of these expected skills. One of the authors was also an instructor for Math 220 for several years and so we relied heavily on his experience in this course in the development of the test. Test items were drawn from a number of sources. Several were adapted from the Field-Tested Learning Assessment Guide (FLAG) (Ridgway et al, 2001), and the remainder were drawn from local precalculus exams, Math 220 exams, or were newly created. Wherever possible, test items were first run in an open-ended form and the most common incorrect answers were used to create distractors in the final multiple-choice version. The initial test was created in Sept 2010 and since then it has been administered in Math 220 a total of 12 times, with 3 substantial revisions and many minor revisions. Table 1 shows the timeline of the development of the test.

Table 1. Timeline of the development of the BPST

Version	Changes	Date of Creation	Dates Administered
V1 – open-ended and multiple choice	Initial version	Sept 2010	Sept 2010 (pre-test), Dec 2010 (post-test), Jan 2011 (pre-test)
V2 - open-ended and multiple choice	Some questions removed, some new questions, and some questions converted to multiple choice	April 2011	April 2011 (post-test)
V3 – fully multiple choice	Some questions removed, some questions converted to multiple choice	May 2011	May 2011 (pre-test), July 2011 (post-test) Sept 2011 (pre-test), Dec 2011 (post-test) Jan 2012 (pre-test), April 2012 (post-test) May 2012 (pre-test), July 2012 (post-test)

In addition to the written answers on the test items, we used student validation interviews (in the form of think-aloud interviews) to assess the quality of the test, as well as we computed standard classical test theory statistics on each item of the test (item difficulty index, item discrimination index, item-to-total correlation, and item characteristic curves). Test items were removed or modified at each stage based on these interviews and statistics.

Results:

The pooled pre- and post-test scores for the current version (V3) of the test are shown in Table 2, as well as the normalized average learning gains (NALG). Note that only students who wrote both the pre- and post-test are included. NALG are computed using matched pre- and post-test scores for each student and are computed using the formula NALG = (<post-test> – <pre-test>)/(1-<pre-test>), where <pre-test> is the mean pre-test score (in %) and <post-test> is the mean post-test score (in %). The NALG therefore is the proportion of the total possible gain, given the particular pre-test mean score.

Table 2. Mean scores on the BPST and normalized average learning gains.

	Math 220 Pretest (N=192)	**Math 220 Post-test (N=192)**	**Normalized Average Learning Gains**	**T-test p-value for M220 pre- and post-test means**
Full test (out of 16)	9.63	11.83	34.58%	0.0006
Precalculus component (out of 4)	2.19	2.61	23.34%	0.1907
Proof skills component (out of 12)	7.44	9.22	39.04%	0.0005

As we can see from the NALG of ~35%, although students are clearly improving on some of the skills in the test, instructors are justified in their impression that students are not learning as much in this course as desired. Some of this is due to poor precalculus skills, but when we examine the test on a question-by-question basis, we see that there are a few proof concepts that students are consistently failing to improve on. As an example, we can consider problem 12 on the test, shown in Figure 1.

Figure 1. Problem 12 from the BPST

For each of the statements below, indicate whether each statement is
(a) always true: true for any choice of the variables
(b) sometimes true: true for some variable choices, but not for all choices, or
(c) never true: not true for any choices of the variables.

12. for real numbers x and y,

$$\sqrt{x^2 + y^2} < x$$

(a) always true
(b) sometimes true
(c) never true

This problem asks students to consider an open sentence and determine if the statement is true for all variable choices, true for only some variable choices or never true for any variable choices. Even at the end of term, only 71% of students correctly answered this problem. For another problem, given three "proofs" of the statement, students are asked to choose the proof that is correct and complete (there are 3 choices, one of which is incorrect, and one is incomplete). Even at the end of our transition-to-proof course, only 60% of students choose the proof that is correct and complete. In addition, the majority of wrong answers choose the incorrect proof that starts by explictly assuming the statement to be proved.

Discussion and Future Directions:

We have found that the BPST can be useful for identifying student difficulties at the start of term, as well as for tracking learning gains in our transition-to-proof course. Indeed, it has illuminated several persistent misconceptions to target in the course. However, further work is needed to establish the validity and reliability of the test, and also to determine if it would be useful to the broader transition-to-proof teaching community. Specifically, the test items need to be validated with a broad demographic of students to ensure they are being interpreted consistently and that wrong answers are chosen for the reasons that we assume. In addition, it is important to get feedback on the test from faculty at other institutions to ensure content validity.

Another possible use for this type of test in the future is to examine the retention of knowledge and skills to future courses. We have recently collected a small amount of data on this, by administering the BPST to students at the end of the Math 220 course, and then again at the start of the first analysis course, Math 320 – Real Variables I. We have a group of 27 students who have completed both of these tests. The mean scores on these tests are shown in the following table.

	Math 220 Pretest (N=27)	**Math 220 Post-test (N=27)**	**Math 320 start-of-term test (N=27)**
Full test (out of 16)	11.70	13.52	13.63
Precalculus component (out of 4)	2.59	2.93	3.00
Proof skills component (out of 12)	9.11	10.59	10.63

As we can see, the mean score in both groups is nearly identical, suggesting that there is very good retention of these skills to Math 320. In addition, these are not just skills that students already had coming into Math 220 because the mean pre-test score for this group is significantly lower ($p<0.001$). Of course, there are many skills beyond these basic ones that are not included in this test that play an important role in learning in a proof-intensive course. It would be another, bigger challenge to identify these and develop a similar instrument to measure them.

Discussion Questions:

- Are there key basic skills missing from our test? Is the set of basic skills different for other types of introductory proof courses?

- What further work is needed to improve the test and establish its validity and reliability? Also, how broadly useful could such a test be (i.e. specific to our institution and course, or more general)?

- This basic proof skills test has been helpful in identifying student difficulties and also in tracking learning, but we would also like to be able to track higher-level proof skills for large numbers (entire classes) in a systematic way. Is it possible to assess higher-level proof skills with a short test? If so, what format could be used for this?

References:

Adams, W. K. & C. E. Wieman. (2010). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education,* 33, 1289-1312.

Ridgway, J., M. Swan and H. Burkhardt. (2001). Assessing mathematical thinking via FLAG. In D. Holton (Ed.) *The Teaching and Learning of Mathematics at University Level* (pp. 423-430). Kluwer Academic Publishers.