Asking the Wrong Questions:  
The Subjective Aspect of Objective Testing  
Thomas D. Malone / Christopher A. Izquierdo

This paper was completed and submitted in partial fulfillment of the Master Teacher Program, a 2-year faculty professional development program conducted by the Center for Faculty Excellence, United States Military Academy, West Point, NY, 2013.

Introduction

When we were charged with the task of drastically redesigning the largest course at USMA, Military Science 100 – Introduction to Warfighting (MS100), which is taught the same term to all fourth class cadets, we knew there would be some tough decisions to make as well as some growing pains. What we didn’t realize at the time was how painful the growth would be that we would endure, nor did we understand how difficult some of those decisions would be. For the sake of this paper, we will not be exploring the depth of our difficulties or the extent of our suffering; however, we will take a closer look at one particular area, the creation of our newly designed assessments and how successful we were at implementing a new approach. We’re going to use this paper to take a look at the questions we included on our assessments—specifically, what went right, what went wrong, and what conclusions we can draw from the lessons learned.

Background of Course Redesign

Before we get into the approach we took for the assessments, it’s appropriate to provide some context and explain why exactly the course needed to be redesigned. The driving force behind the course redesign was a department-wide effort to realign the various components of a cadet’s military education. These components include input and contributions from three different divisions within the Department of Military Instruction (DMI): Military Science, Military Training, and Accessions. You have the classroom component provided by Military Science, the military training provided by Military Training, and the branch education provided by Accessions. Based on feedback received through formal After Action Reviews (AARs) and the end of course critiques, it was clear that cadets viewed the different components of their military education as distinct and somewhat disjointed. Any sort of linkage between summer training and the cadet’s classroom instruction during the academic year either didn’t exist or we weren’t doing an adequate job of bridging the gap and demonstrating that connection between what is taught in the classroom and implemented in the field environment.

The cadets are required to take three different blocks of classroom instruction from the Military Science Division, MS 100, 200, and 300. In previous years, all three of these blocks more or less resembled one another, with the main difference being an escalation of complication and difficulty from one year to the next. For the redesigned course, we really wanted to redefine what we viewed as the focus of the course—shifting from what cadets were capable of learning to the utility of what they’re being taught. This new philosophy also aligned well with our renewed emphasis on synchronizing the classroom and field activities for the cadets.

The cadets that we primarily deal with for MS100 are fourth class cadets. These students endured Cadet Basic Training (CBT) the previous summer as new cadets, enjoyed a one term break from military training in the fall, and are now mentally preparing themselves to participate in Cadet Field Training (CFT) as team leaders in the summer that follows. Our revamped course was designed to make all of these events seem like one continuous, coherent block of instruction rather than a few disjointed blocks of instruction whose only similarities were that they dealt with military topics. By achieving this we would help prepare them for their upcoming summer training as well as MS200, which they would all take the following academic year.
**Testing Method Shift**

As mentioned, the three Military Science courses (100, 200, 300) have historically been very similar. However, their similarities didn’t end with content—the approach to teaching as well as the way tests were administered and graded was extremely similar as well. These courses weren’t filled with lectures and they weren’t full of facts that needed to be memorized. Rather, they were overflowing with Tactical Decision Exercises (TDEs) and questions that had numerous correct answers depending on how well a cadet was able to argue their position. These were enjoyable classes to teach; however, they had almost nothing to do with those things that fourth class cadets needed to know at that stage of their development and had virtually no relationship to any aspect of their summer training.

With the redesign, we quickly realized that the approach and testing procedures we had used before would not work for our purposes and that some drastic changes would need to be made, even if that involved a shift away from the extremely subjective grading practices that our department and its members had finally begun to embrace after years of resistance. There were four basic areas that we decided to focus on that would effectively bridge the gap between CBT and CFT: Marksmanship, Land Navigation, Tactical Communications, and First Aid. To put it in the terms of the Soldier’s Manual of Common Tasks, we were going to shift our focus to the basic Soldier skills of “Shoot, Move, Communicate, and Survive.” With this new approach, the subjective grading approach would not be acceptable. We would no longer be dealing in shades of gray, but black and white—with the shift of our focus we were now dealing with material where there was clearly a right answer, a wrong answer, and very little room for interpretation. Therefore, it seemed extremely logical to shift our approach to objective-based testing. We were now dealing with basic fundamentals that needed to be studied and committed to memory.

Reflecting on previous blocks of instruction in the Master Teacher Program, we recalled seeing Bloom’s Taxonomy (See Figure 1). In the past, MS100 had focused on Comprehension and Application, which we believed was the wrong approach. We were going to pull back on the proverbial reins and focus their efforts on Knowledge—building up the foundational information that cadets had been failing to display as they completed their first year at the academy. This sort of foundational knowledge is custom built for objective tests that include true/false, short answer, and multiple choice questions. Now that we had identified the types of questions we wanted to ask our cadets, it was time for us to create the assessments for each of our four foundational modules.

**Question Creation & Analysis**

We analyzed the stats for each of these four modules and annotated what we determined were our outliers, whether good or bad. Refer to Figures 2-5 for graphs of the statistics from these assessments. On these graphs, we’ve annotated two types of questions—those with a high proportion of correct answers (Green Square) and those with an abnormally large number of incorrect answers (Red Square). Both ends of the spectrum were determined by the standard USMA grading scale. We deemed too high of a score to be 95% or greater since we believe a well constructed, challenging question should generate more incorrect responses; on the other end of the spectrum, we decided if 30% or more of our students missed the same question that’d be the equivalent of a D or worse average on that particular question and it would warrant a second look.
Question Outliers

Through all four assessments, we identified a total of 29 question outliers—15 at the high end of the spectrum and 14 at the low end of the spectrum. We will focus the majority of this brief discussion on the latter of the two categories. Those questions at the high end of the spectrum were pretty straight forward as far as our determination of why we believe they were so easily answered. Simply put, the answers to these questions were far too obvious and the questions weren’t difficult enough to truly challenge the cadets. In the future we would omit these questions or modify them in order to make them more taxing.

As for the questions on the low end of the spectrum, they seemed to fall into a few different categories, but there were two characteristics that seemed to be fairly common for most of the 14 questions—the format of the questions tends to be short answer and multiple part questions seem to be more difficult. The idea that a multiple part question might be more difficult than those questions with one part seems relatively common sense. However, why might short answer questions appear so frequently on this list? When we were developing these assessments, our questions were divided almost equally between short answer, multiple choice, and true/false. The reason these short answer questions appear to be missed at a rate higher than what we’d desire seems to be two-fold in our minds, 1. On the student’s end, unlike with the other question-types, a short answer does not provide them with an answer from which to choose, and 2. On the instructor’s end, short answer questions have much more room for error and it’s difficult to create a clear, truly objective question.

For the students, when they are given multiple choice or true/false questions they are provided the answer; however, when dealing with short answer questions they must provide this information on their own. The question itself may not be exceptionally challenging, but when faced with a time crunch and the added pressure of an assessment being graded, answers to even the simplest of questions can elude a student’s ability to recall the information. For the instructors, it proved to be a challenge to create short answer questions that truly only had one answer. As we were creating, revising, and administering our tests, it was not uncommon for someone (whether it was an instructor or a student) to come up with an “off the wall” answer that we hadn’t thought of before, but could arguably be justified as a correct response.

Beyond these two common characteristics, there was one additional thing worth pointing out. On three of the four assessments we had what we described as a “tricky” question that resulted in a lot of students missing that specific question. While it wasn’t our intention when we wrote the question to trick our students it was clear, based on their responses, that this is what ultimately happened. Please refer to Figure 6 for additional information related to the outlying questions.

Conclusion

When we embarked on this journey they call course directing, we never thought it’d be this challenging—better put, we knew it would be challenging, but we didn’t know we would be so challenged by those things that proved to be so problematic, such as the semantics of constructing a specific question. We realized the value, even for objective tests, of a well thought out grading rubric. The tests are not inherently easier because there’s supposed to be one correct answer. This may lessen the burden of grading for instructors, but ultimately it increases the difficulty of creation ten-fold. If you’re interested in conducting any sort of test analytics similar to what we did this term, you’d be best served creating questions that have one single answer and not multiple parts as this makes your statistics more difficult to analyze. As you are creating and vetting your assessments, pay close attention to the short answer questions and try to war-game any other possible answers that may be deemed acceptable.
While we initially created an acceptable range of between 70-95% for correct responses, the true goal should be between 75-90%. If every question garnered a correct response rate that fell within that range, it would set the stage for equitable scores and opportunities across the board. Scores within that range would likely indicate that no students are precluded from earning a perfect score, while at the same time no students are being punished for poorly written or overly difficult questions. There’s nothing wrong with a question being difficult and yielding a low percentage of correct responses; however, this shouldn’t be by accident, it should be by design.
<table>
<thead>
<tr>
<th>Q#</th>
<th>Type</th>
<th>Subject</th>
<th>Reason</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>T/F</td>
<td>&quot;Functions Check&quot; and SAW Disassembly</td>
<td>Tricky</td>
<td>717</td>
<td>313</td>
</tr>
<tr>
<td>M11</td>
<td>SA</td>
<td>Identify Weapon Part (Receiver Assembly)</td>
<td>Difficult</td>
<td>742</td>
<td>348</td>
</tr>
<tr>
<td>L1</td>
<td>SA</td>
<td>Identify Map Colors</td>
<td>Multiple Parts</td>
<td>757</td>
<td>377</td>
</tr>
<tr>
<td>L7</td>
<td>SA</td>
<td>Terrain Features (Draw)</td>
<td>Difficult</td>
<td>735</td>
<td>399</td>
</tr>
<tr>
<td>L12</td>
<td>SA</td>
<td>Curved Line Distance</td>
<td>Difficult</td>
<td>465</td>
<td>669</td>
</tr>
<tr>
<td>L20</td>
<td>MC</td>
<td>Rate of Travel Equation</td>
<td>Tricky</td>
<td>764</td>
<td>370</td>
</tr>
<tr>
<td>L25</td>
<td>SA</td>
<td>Resection</td>
<td>Difficult</td>
<td>676</td>
<td>458</td>
</tr>
<tr>
<td>T4</td>
<td>MC</td>
<td>Radio Modulation</td>
<td>Difficult</td>
<td>673</td>
<td>317</td>
</tr>
<tr>
<td>T9</td>
<td>T/F</td>
<td>Pro-Words</td>
<td>Tricky</td>
<td>332</td>
<td>658</td>
</tr>
<tr>
<td>T10</td>
<td>SA</td>
<td>9-Line MEDEVAC</td>
<td>Difficult/Multiple Parts</td>
<td>532</td>
<td>458</td>
</tr>
<tr>
<td>T12</td>
<td>SA</td>
<td>SALUTE Report</td>
<td>Difficult/Multiple Parts</td>
<td>563</td>
<td>427</td>
</tr>
<tr>
<td>F12</td>
<td>SA</td>
<td>Three Major Causes of Combat Death</td>
<td>Multiple Parts</td>
<td>635</td>
<td>366</td>
</tr>
<tr>
<td>F18</td>
<td>SA</td>
<td>NPA Equipment</td>
<td>Difficult/Analysis</td>
<td>683</td>
<td>318</td>
</tr>
<tr>
<td>F19</td>
<td>SA</td>
<td>Tourniquet Usage</td>
<td>Difficult/Analysis</td>
<td>604</td>
<td>397</td>
</tr>
<tr>
<td>M1</td>
<td>MC</td>
<td>Four Fundamentals of Safety</td>
<td>High Emphasis</td>
<td>1041</td>
<td>49</td>
</tr>
<tr>
<td>M2</td>
<td>SA</td>
<td>Four Fundamentals of Marksmanship</td>
<td>High Emphasis</td>
<td>1039</td>
<td>51</td>
</tr>
<tr>
<td>M6</td>
<td>MC</td>
<td>Max Effective Range (M240B)</td>
<td>Very Easy</td>
<td>1060</td>
<td>30</td>
</tr>
<tr>
<td>M10</td>
<td>SA</td>
<td>Identify Weapon Part (Trigger Mechanism)</td>
<td>Obvious</td>
<td>1054</td>
<td>36</td>
</tr>
<tr>
<td>L5</td>
<td>T/F</td>
<td>Two Letter Grid Square Identifier</td>
<td>Very Easy</td>
<td>1096</td>
<td>38</td>
</tr>
<tr>
<td>L6</td>
<td>SA</td>
<td>Terrain Features (Hilltop)</td>
<td>Obvious</td>
<td>1101</td>
<td>33</td>
</tr>
<tr>
<td>L13</td>
<td>MC</td>
<td>Degrees vs. Mils</td>
<td>Very Easy</td>
<td>1099</td>
<td>35</td>
</tr>
<tr>
<td>L14</td>
<td>SA</td>
<td>Converting Magnetic to Grid</td>
<td>Very Easy</td>
<td>1083</td>
<td>51</td>
</tr>
<tr>
<td>T2</td>
<td>MC</td>
<td>ASIP Characteristics</td>
<td>Very Easy</td>
<td>943</td>
<td>47</td>
</tr>
<tr>
<td>T8</td>
<td>T/F</td>
<td>Radio Protocol</td>
<td>Obvious</td>
<td>957</td>
<td>33</td>
</tr>
<tr>
<td>T11</td>
<td>MC</td>
<td>MEDEVAC Support</td>
<td>Very Easy</td>
<td>974</td>
<td>16</td>
</tr>
<tr>
<td>F1</td>
<td>SA</td>
<td>Heat Exhaustion</td>
<td>Obvious</td>
<td>963</td>
<td>38</td>
</tr>
<tr>
<td>F2</td>
<td>MC</td>
<td>Unacceptable First Aid</td>
<td>Obvious</td>
<td>962</td>
<td>39</td>
</tr>
<tr>
<td>F4</td>
<td>T/F</td>
<td>Heat Cramps</td>
<td>Obvious</td>
<td>988</td>
<td>13</td>
</tr>
<tr>
<td>F22</td>
<td>T/F</td>
<td>Tourniquet</td>
<td>Obvious</td>
<td>986</td>
<td>18</td>
</tr>
</tbody>
</table>

Figure 6 - Question Outliers