2014

# The Importance of Formative Assessment in Science and Engineering Ethics Education: Some Evidence and Practical Advice

Matthew W. Keefer
*University of Missouri-St. Louis*

Sara E. Wilson
*University of Kansas*

Harry Dankowicz
*University of Illinois at Urbana-Champaign*

Michael C. Loui
*Purdue University*, mloui@purdue.edu



Keefer, Matthew W.; Wilson, Sara E.; Dankowicz, Harry; and Loui, Michael C., "The Importance of Formative Assessment in Science and Engineering Ethics Education: Some Evidence and Practical Advice" (2014). *School of Engineering Education Faculty Publications.* Paper 10.
http://dx.doi.org/10.1007/s11948-013-9428-5

# The Importance of Formative Assessment in Science and Engineering Ethics Education: Some Evidence and Practical Advice

Matthew W. Keefer
Department of Educational Psychology, Research and Evaluation, University of Missouri–St. Louis, St. Louis, MO 63121, USA

Sara E. Wilson
Department of Mechanical Engineering, University of Kansas, Lawrence, KS, USA

Harry Dankowicz
Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Michael C. Loui
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

December 8, 2012

**Abstract** Recent research in ethics education shows a potentially problematic variation in content, curricular materials, and instruction. While ethics instruction is now widespread, studies have identified significant variation in both the goals and methods of ethics education, leaving researchers to conclude that many approaches may be inappropriately paired with goals that are unachievable. This paper speaks to these concerns by demonstrating the importance of aligning classroom-based assessments to clear ethical learning objectives in order to help students and instructors track their progress toward meeting those objectives. Two studies at two different universities demonstrate the usefulness of classroom-based, formative assessments for improving the quality of students' case responses in computational modeling and research ethics.

This paper will appear in *Science and Engineering Ethics*, DOI: 10.1007/s11948-013-9428-5

## 1. Introduction:

Recent research in ethics education shows both encouraging and potentially problematic trends. While studies show a positive increase in course and program offerings, they also demonstrate a large variation in content, curricular materials, and instruction (DuBois et al., 2010; Kalichman & Plemmons, 2007; Lehmann, Kasoff, Koch, & Federman, 2004). For example, DuBois et al. (2010) conducted an extensive survey of mandated RCR instruction and found that while RCR instruction is widespread (97% of those surveyed), there is "no unified approach (p. 109)"; there exists rather a "significant variation in scope, content, and approaches to RCR instruction (p. 110)". These authors report that programs "lack a coherent plan for RCR instruction." Furthermore, since 82% of instructors claim to be using "original" curricular materials, this makes the challenge of assessing the value of these materials more daunting. In addition, Kalichman and Plemmons (2007) found great variation in the learning goals chosen in RCR ethics education. The ethics instructors they surveyed identified no less that 50 distinct goals. While it could be argued that having a large variety of instructional approaches can provide some benefits, especially in new areas, the observed variability in instructional goals within the same content areas raises the specter of significant problems with curricular alignment. According to Kalichman and Plemmons (2007), "identifying effective strategies for RCR education depends on first defining measurable outcomes based on well-defined goals. The findings of this study suggest a lack of consensus about those goals (p. 846)." The educational implications are more bluntly expressed in Kalichman (2007) as, "the stated goals and purpose of RCR education are diverse, inconsistent, and sometimes not feasible (p. 870)." Similar concerns have been identified in other core areas. Lehmann et al. (2004) report that while medical ethics course offerings have increased, "significant variation in the content, method, and timing of ethics education suggests consensus about curricular content and pedagogic methods remains lacking (p.682)".

These findings suggest that while the numbers of ethics courses and programs have risen, research has uncovered serious problems with the present design of ethics education. We suspect that part of the reason for these problems is that ethics education is the victim of its own success. With a significant increase in more required ethics courses over the past two decades, there are also more curricular resources and materials available to support instruction, i.e., cases, commentaries, videos, role-playing scenarios, micro-insertions, etc. While it is salutary that more professional ethics instruction and instructional materials are available, problems arise as these materials are typically provided with little advice on how to use them. Furthermore, many instructors responsible for developing their own courses come from a wide variety of disciplines and are often teaching a subject that is not their primary area of expertise. To make matters worse, it is often not recognized that the learning objectives in professional ethics education are ambitious. Realistic ethical problems are what cognitive scientists refer to as "ill-structured problems", because there is no clearly specified goal, usually incomplete information, and multiple possible solution paths. Since a simple response is not likely the best option, students must investigate the problem, seek relevant information, consider alternative actions, and evaluate short and long-term consequences. These problems are instructionally challenging because good student responses can lead in quite different directions, providing emphases on a diversity of values and issues that are difficult to predict. It is then easy to recognize why quality ethics curricula can make the assessment of student learning a challenge.

These considerations also help explain why many ethics courses are not adequately designed or assessed. For example, the findings described above fit our experience that while most professional ethics courses spend a considerable amount of time discussing ethics cases and perhaps have students submit written responses to case examples, more often than not these responses are either not assessed, or graded simply as pass/fail. Below we will offer a comparison of two different instructional strategies for assessing students' responses to complex cases in the field of computational modeling and research ethics. The comparison will exemplify both a common problem as well as a potential solution to the assessment of an important component of ethics education.

To summarize, while we acknowledge findings showing that more consistency in learning objectives and instructional activities is needed, it is also necessary that ethics instructors understand how to align instructional activities to learning objectives and to assess whether these activities are effective. The learning objectives in this case are *(1) to help students to be sensitive to the ethical problems they may face in their professional practice and (2) to develop a good plan of action in response to them*. The goal of this paper is to demonstrate the importance of aligning some professional ethics learning outcomes with instructional activities and, most important, ensuring that they are appropriately assessed. When learning activities are not properly assessed, students will be uncertain of what they are supposed to learn and instructors will lack the opportunity to make informed adjustments to their instructional activities. Some recent evidence for this claim will be presented below, but first it is important to understand some of the different approaches to the assessment of student learning.

## 2. Understanding the difference between classroom and standardized assessments

Over the past couple of decades, the role of assessment in ethics education has been motivated by efforts to justify the inclusion of ethics in the science and engineering curriculum. This type of assessment characteristically employs standardized assessments. In comparison with classroom-based assessment, standardized assessments test general skills and are administered under controlled test-taking conditions. The aim of these assessments is to determine whether a program or course had a positive effect on some variable (e.g., moral reasoning), usually in comparison to a baseline (e.g., pre-test). Employed in this manner they can provide some independent evaluation of whether the course or program of instruction is worthwhile.

The most common example of a formal measure used in ethics education is James Rest's Defining Issues Test (DIT); the current version is Defining Issues Test-2 (DIT2). For example, the DIT2 provides a valid measure of level of moral reasoning based on Kohlberg's stage theory, and has been used successfully to assess professional ethics instruction (Loui, 2006; Bebeau, 1995). The advantage of measures such as these is that they have been extensively tested over a wide population of subjects and have been proven to be valid and reliable.

While standardized ethics assessments have been important for justifying the inclusion of ethics in science and education curriculum, they are not otherwise practical. Why? From the perspective of the classroom, these assessments can be costly and time consuming. They also do not tell much about what is learned in any specific course. The administration of standard assessment tests raises an ethical issue of whether students should be required to take up valuable instructional time to complete tests unrelated to their curriculum Davis and Feinerman, (2010). It also brings up the related concern of whether the instructor should be using instructional time administering assessments if they do not help them know what they need to know *as instructors*, i.e., whether their students are learning what they expect them to be learning. The focus of this paper will be on what are referred to as *formative* assessments: assessments that are usually designed by instructor to assess their students' learning, as well as inform instructional activities and design.

### 2.1 Understanding the importance of formative assessments

When designing instruction, educational psychologists distinguish between *formative* and *summative* assessments Angelo and Cross (1993). Summative assessments focus on measuring the acquired learning after completion of instructional activities. The role of formative assessments, in contrast, is on how well students understand what they are presently being taught. The use of formative assessments will be emphasized in the discussion below.

The effectiveness of formative assessment has long been acknowledged in educational research (Resnick, 1987; Angelo & Cross, 1993; Black and Williams 1998; Bransford, Brown, & Cocking, 2008). For example, Black and Wiliam (1998) completed an influential review of research on formative assessment practices encompassing P-20 (i.e., kindergarten to college), over a range of content areas and in various countries. They reported gains in student achievement with effect sizes between 0.4 and 0.7.

Formative assessments are primarily of interest to the instructor and the students during the course of instruction. For this reason, formative assessments are most useful if provided to students (in the form of checklists, rubrics or scoring guides, or models across various performance levels) prior the completion of course assignments. Formative assessment thus plays an essential role in instructional design. Its use is an area that clearly needs more attention in ethics education (Keefer & Davis, 2012). Here we report a pilot study and a follow up study that provides evidence for this claim.

## 3. Pilot Study

The studies reported below were designed as part of an NSF-funded Ethics Education in Science and Engineering research grant, whose goal is to develop instructional materials to teach the standards specific to computational modeling and research (Kijowski, 2010; Kijowski, Dankowicz and Loui; 2011). As part of this effort, we developed a series of case scenarios involving issues in computational modeling and research ethics, and tested the effectiveness of two of these case scenarios in four graduate-level engineering courses. The test design and characteristics of our students will be described below.

### Method

For both of the studies reported we received the approval of the Institutional Review Board at three different institutions: the University of Illinois at Urbana-Champaign, University of Missouri – St Louis, and the University of Kansas. Students completed consent forms, which the instructors did not access. All of the data analysis was completed at the University of Missouri – St Louis.

In the fall 2010 semester, Sara Wilson at the University of Kansas piloted our first case scenario by asking students in an RCR course for first year graduate students in Bioengineering and Mechanical Engineering (N=7), to provide a written commentary as homework and then discuss the case and their responses in class. The students were informed that their written responses were to be graded as either pass/fail. The case consisted of two parts, the first part of this case appears in Table 1.

**Table 1.**
**The Mentorship Matrix**

**Part 1**

Dr. Smith is a leading researcher in the development of novel pharmaceutical nanoparticles and the development of new protein-based pharmaceuticals. His lab is an exciting collaborative environment where some graduate research assistants work on developing computational models of protein structure interactions, other students work on using this modeling software to come up with new proteins that might make promising new pharmaceuticals, and still more graduate students work on creating and testing these proteins. In this collaborative environment, students help each other, and students often rely on the work of other students. There is an expectation in the lab that students' work may be shared and applied in other projects. Dr. Smith is a popular thesis advisor with many students. Most students in his lab are able to get jobs in the pharmaceutical industry, and most of the research in the Smith lab is funded by pharmaceutical companies looking to develop the next new wonder drug.

Nanoparticles are intriguing in that they offer a way to package a pharmaceutical particle to control the release profile of the drug. Dr. Smith came up with an idea for modeling the interaction of the nanoparticle polymer components and protein - based pharmaceuticals mathematically. Matrix Pharmaceuticals was interested in funding Dr. Smith's lab to develop this idea further. He hired Mr. Anderson as a graduate research assistant to create the software that implements this mathematics in a computational model of the nanoparticles. Mr. Anderson is a bright and talented programmer with an undergraduate degree in computer science, but he is not the most organized. For his MS thesis project with Dr. Smith, he created a code that solves these mathematical formulations using a novel nonlinear optimization algorithm. This formulation is particularly fast for systems with sparse characteristics, but can be problematic for non-smooth objective functions. As Mr. Anderson wrote his code, he was in a hurry as he has a job offer from Zion Pharmaceuticals in two months. He didn't bother to comment the code while writing it, since he believed he would have time to do that later when he was writing the thesis. The code is all in one large file rather than broken down into modules, making it very hard for anyone other than Mr. Anderson to follow. Once he wrote the code, he was able to validate his code for the nanoparticle polymer-protein

> formulations the lab had already created. The software correctly predicted their size and material properties.
>
> While he was finishing his work in the lab and preparing for his new job, Mr. Anderson assured Dr. Smith that he would be able to get the code commented and thesis written within the first month of working at Zion. However, when Mr. Anderson got to the job, he found the job to be very demanding and, consequently, he has not made the progress he had planned.
>
> Matrix Pharmaceuticals has asked Dr. Smith to design a new nanoparticle-vaccine formulation for the treatment of HIV in the hope that such a formulation might hold the key to a potential vaccine for HIV. Dr. Smith thinks that Anderson's code would be very useful for this project and is considering asking another student to begin working on the project.

Students' responses using the instructional procedure described above were disappointing both in regard to being all too brief and often incomplete in identifying significant ethical issues. In the following semester, it was decided that the students would be provided with a formative assessment – referred to as a Decision Procedure Checklist (DPC). The DPC was designed to provide students with practical guidance in considering four important components of realistic ethical problems. In the terminology of educational theory, the DPC is a *formative* assessment tool that helps students focus their learning on what is presently being taught. In terms of ethical reasoning, the DPC helps students focus on four important learning objectives that pertain to complex ethical problems. The DPC prompts students to (1) identify ethical issues and professional responsibilities, (2) identify additional important information (investigate the problem), (3) consider alternative courses of action in response to the case, and (4) consider the long and short-term consequences of proposed solutions. The development of this checklist was based on findings from previous research on how experienced ethicists respond to realistic ethical cases (Keefer & Ashley, 2001).

**Table 2. Decision Procedure Checklist (DPC)**

- Ethical issues/professional responsibilities identified
    - Have the primary and secondary stakeholders been identified? Stakeholders can include, i.e., individuals, groups, societies, companies, etc.
    - Have the ethical issues been identified and how they relate to various stakeholders been considered?
- Additional information Identified
    - Has additional useful knowledge or information concerning the problem been identified?
    - Are any additional resources identified that could help in developing a solution to the problem?
    - Have actions been taken that could provide additional useful information or provide additional resources?
- Consideration of the actions taken in response to the case
    - How well do recommended actions address the concerns of primary and secondary stakeholders?
    - How well do the recommended actions address the ethical issues identified?
    - Are there any creative "middle way" courses of action that can address more than one ethical issue?

- Consideration of long and short-term consequences of proposed solutions
    - Is there consideration of how the proposed solution might affect the stakeholders in the problem over time?
    - Have any morally significant longer-term consequences of the proposed solution been considered (including possible accidents, misuses, etc.)

In a second offering of the same course (N=10), students were again asked to provide a written commentary of the same case prior to class, but were explicitly asked to use a Decision Procedure Checklist (DPC) to guide their responses. They were instructed that their responses would be assessed and graded relative to the checklist.

## Results

We believed that the student responses in the second course showed marked improvement. To test this belief, a graduate research assistant (GRA) scored each of the student responses for both of the courses using a Decision Procedure Scoring Guide (DPSG). The DPSG can be considered a *summative* assessment that applies a four-point graded scale for each of the 4 components identified in the DPC *formative* assessment - i.e., Less Proficient 1, Proficient 2, More Proficient 3, Expert 4 (see Table 3). The GRA was blind to the research hypotheses for both studies and, in addition, from a different institution than any of the students tested.

**Table 3.**
**Decision Procedure Scoring Guide (DPSG)**

**Identify ethical issues/professional responsibilities:**

| *Expert* | *More Proficient* | *Proficient* | *Less Proficient* |
|---|---|---|---|
| Identify all relevant ethical issues in the case and how they relate to professional responsibilities. Identify and track concerns of primary & secondary stakeholders. | Identify more than one ethical issue and/or a professional responsibility. Identify and track concerns of the primary stakeholders. | Identify a key ethical issue and/or a professional responsibility. Identify the concerns of a primary stakeholder. | Only a single ethical issue is identified. Or, the problem is not considered to have an ethical dimension. |

**Identify additional information (investigate the problem)**

| *Expert* | *More Proficient* | *Proficient* | *Less Proficient* |
|---|---|---|---|
| Recognize and appropriately use resources that support ethical action (or that failed to). Identify all additional knowledge or information that | Recognize some resources that might support ethical action. Identify some additional knowledge or information that is useful and is | Some potentially useful information is considered but may not effectively be incorporated into the solution. Additional resources are not | Additional resources and information are not recognized or incorporated into the proposed solution. |

| | | | |
|---|---|---|---|
| might useful to know and identify appropriate action. | incorporated into the proposed solution. | considered. | |

**Consider alternative courses of action in response to the case**

| *Expert* | *More Proficient* | *Proficient* | *Less Proficient* |
|---|---|---|---|
| The recommended course of action addresses several ethical issues simultaneously. The solution also addresses and tracks the concerns of all relevant stakeholders. | The recommended course of action addresses more than a single ethical issue effectively. The solution addresses the concerns of more than a single stakeholder. | The recommended course of action addresses a single ethical issue effectively. The solution addresses the concerns of a single stakeholder. | The recommended course of action does not address a key ethical issue effectively. The solution does not adequately address the concerns of any stakeholders. |

**Consider the long and short-term consequences of proposed solutions**

| *Expert* | *More Proficient* | *Proficient* | *Less Proficient* |
|---|---|---|---|
| The solution anticipates all morally significant short and longer-term consequences of actions. Morally significant alternative actions are considered anticipating changing circumstances or events. | The solution recognizes some morally significant short and longer-term consequences of actions. Some alternative actions are considered anticipating changing circumstances or events. | The solution recognizes some morally significant consequences of actions. | The solution does not recognize morally significant consequences of actions. |

In support of our hypothesis, the grand mean scores for the two courses were 1.64 and 2.93 respectively. In other words, the difference between the two means was statistically significant, $t(15) = -4.107$, $p < .001$ (Keefer & Wilson, 2011). Assumptions of homogeneity and normality were tested and homogeneity was violated while normality was not. While these results were very encouraging, we could not be sure the formative assessment was the source of the improvement since we were comparing students from two different courses without the benefit of any baseline data.

## 4. Method

In the fall of 2011, we decided to implement a more controlled comparative study using two different courses in computational mechanics and in finite-element analysis, respectively, at the University of Illinois at Urbana-Champaign. We employed a pre- and post-test strategy. For each course, students wrote a response to one case before an instructional session as a pre-test. The instructional class session was conducted by Harry Dankowicz, in both courses, the students wrote a response to a different case after each session as a post-test. Both cases consisted of two separate parts to which the students were asked to respond. The students in course I received the DPC prior to completing both case responses, whereas the students in course II received the DPC only for the second case. The DPC served as the intervention (i.e., formative assessment) with the DPSG as the dependent measure. The two different cases were presented in the same order in both courses. The students were told their responses would not be graded.

**Results**

All student responses were scored blindly by the graduate research assistant. In addition, the GRA did not know which student responses were submitted before the class session and which were submitted afterward. We conducted a paired-sample t-test to compare the mean scores of pre- and post-test, in order to see if the difference was significantly different from zero. In course I (N=36) the mean scores were 2.61 and 2.51 respectively and the difference was not statistically significant. By contrast, in course II (N=47) the mean scores were 1.81 and 2.23 and the mean difference was significant with a large effect size $t(46) = -5.63$, $p < .001$, $d = .90$. Assumptions of homogeneity and normality were tested and not violated. As the results in course I were not significant (e.g., the students' responses to the first case were scored higher than the second), the statistically significant effect in course II is unlikely explained by variation in difficulty between cases. However, the overall mean scores for course II are appreciably lower than the course I scores, indicating that the quality of the case reasoning was lower in course II. This somewhat surprising result may perhaps be due to variation in the students enrolled in the classes. Both were available for undergraduate (elective) or graduate credit and students were not differentiated based on their enrollment status - though a significant proportion of graduate students was expected given the role that these courses play in the Computational Science and Engineering Graduate Option Program at the University of Illinois at Urbana-Champaign. It is also important to note that, due to practical limitations, the instructor was allotted less instructional time in the second class session than the first. This suggests the formative assessment (i.e., DPC) may provide a more significant effect than instruction, which would further recommend their inclusion in brief ethics modules or insertions such as these. The difference in instructional time cannot explain the significant differences between the pre-test scores, however.

To reduce concern that overall differences in scores between courses might be due to coding variation, we conducted a word count for each student response to both cases in both classes. We hypothesized that if the lower scores in course II were due to the students' effort or ability as our coding indicates, there should be a significantly lower mean word count for both case responses in course II compared with course I. This hypothesis was supported, as the mean word count in course II was 248 (case one) and 276 (case two) as compared with 496 (case one) and 366 (case two) in course I.

The importance of the statistical findings is of course to confirm the improvement in students' thinking pre-test to post-test. It is also useful to examine more qualitative effects. When students are prompted by the formative assessment (i.e., the intervention) they appear to focus their thinking on the more important and challenging aspects of complex problems, as we think the example in Table 4 from the University of Kansas sample demonstrates.

**Table 4.**
**Student Response to The Mentorship Matrix**

**Part 1**

Ultimately, Dr. Smith should coordinate and communicate with Mr. Anderson about the situation. Chances are that Mr. Anderson would be okay with his code being used in further research. Each university, however, has special clauses within contracts that dictate the use of intellectual property developed or designed on university grounds or with university funds. The university has a role in the dissemination of intellectual property (in the case of University of Kansas, 1/3 ownership at least). Dr. Smith and Mr. Anderson are stakeholders in the case and conflict should be resolved between them before major steps are taken.

More information is needed in order to understand the capabilities of the program. For instance, if the regressed (or predicted) data that the program has produced will constitute a completely different class of carriers than the ones suggested, further validation studies should be implemented to understand the robustness of the program. This would avoid any unnecessary costs associated from Matrix Pharmaceuticals which Dr. Smith has a professional obligation to. Note, that if not advertised under false pretenses, Dr. Smith has no legal requirements to produce working data from the program, but such a job would not be the best ethical or professional choice.

Depending upon the contract or negotiation, Mr. Anderson may or may not have an obligation to go back and finish his work. Chances are that he has no legal requirement to do so unless an oral contract was negotiated per terms of his graduation. Regardless, the amount of time for Mr. Anderson to comment the code would be significantly smaller than if any other third party were to come in to decipher the code. Once again, on the basis of professionalism and common courtesy, Mr. Anderson should fulfill the duties which he promised.

Based off my interpretation of the parties and situation at hand, this problem could easily be averted. It seems that Mr. Anderson has little or no interest in the program in question and probably would not care that other people utilized his work. After a phone call or email, this can easily and quickly resolved so that Dr. Smith may make progress on the program in order to help out Matrix Pharmaceuticals and potentially HIV patients.

As this response shows, the student effectively attends to the complexity of the case by considering alternative actions that are contingent on important information and issues (e.g., contracts, intellectual property). Furthermore, the student identifies several areas where additional information and knowledge would be useful; considers what must be done to acquire it; and what should happen depending on the outcome. There is a useful mix of identifying both practical and ethical considerations that take into account both short and long-term ethical consequences, a characteristic that is often observed in the responses of experienced ethicists (Keefer & Ashley, 2001; Harris, Pritchard, & Rabins, 2009).

## 5. Educational Implications:

The educational benefits reported here should make clear the positive role that formative assessment can have in both the creation and improvement of instructional designs. An obvious benefit of quality formative assessments is that they should provide relevant feedback to the instructor that will help adjust future instruction to better meet the learning goals. In this case, examining the students' aggregate four component scores of the rubric (and checklist) could provide valuable feedback for where future instructional resources might be allocated. For example, the instructor might find that students are doing a better job at identifying ethical issues and the ethical resources they need while not paying enough attention to the some of the short and long term consequences in their case responses.

The benefits can also accrue with continued use. To provide one example, once instructors have gathered a range of student responses applying a formative assessment (with permission), these can provide new students with models of the varying performance levels. New instructional activities can then be designed using the formative assessment by including student responses 'anchored' to the assessment's indicators. For example, assuming there are enough cases, students in small groups can practice scoring the sample student responses to the scoring guide's indicators as an ungraded activity. Creating classroom activities like these provide students and instructors with a continuous "feedback loop" Angelo and Cross (1993). This helps students self-assess the progress of their learning while instructors are able to track how effective their learning designs are at fostering the targeted learning goals.

A critical point could be made that providing students with a rubric is equivalent to 'telling them what you want them to do' and so may not provide the best measure of students' ability to generate a strong case analysis, or transfer that knowledge to their subsequent professional contexts.  While this study cannot address the latter issue directly, we would respond that what one 'wants the students to do' is to engage in learning activities that are appropriately aligned with worthy learning goals. The learning objectives in this case are to help students be sensitive to the ethical problems they may face in their professional practice and to develop a good and systematic plan of action in response to them. The method we used to develop the computational modeling problems is discussed in a previous study published in this journal (Kijowski, Dankowicz, & Loui, 2011) but it included extensive interviews with ethicists and experts in the field, hence these problems are likely to represent the ones these students might actually face. While this is no guarantee, it improves considerably the chances that students will transfer their learning. Similarly, we believe the improvement in our students' responses (e.g., the sample student response) results from the alignment of our formative assessment (i.e., rubric) with the goal of helping students develop a good plan in response to realistic and complex professional problems. The case for our rubric's alignment has more than prima facie validity as its components were derived from empirical research that provided a systematic comparison of novice and experienced ethicist's case reasoning using similarly realistic and complex problems (Keefer & Ashley, 2001; Harris, Pritchard, & Rabins, 1995). In terms of instructional best practice, it should also be noted that when using a rubric or scoring guide it is important to supply it 'up front' so that students will know how they will be evaluated prior to grading and not be subject to what they experience as an unpleasant 'surprise' (Stevens, & Levi, 2013; p. 50).

Finally, we note an unfortunate use of the word "all" in the characterization of "Expert" proficiency in the DPSG. To avoid giving the impression that unreasonably high expectations are required for this level of proficiency, we recommend (and in future use of the DPSG, plan to) replace "all" with "the" throughout.

**6. Conclusion**

As we stated at the outset, our goal is to demonstrate the importance of aligning learning outcomes with instructional activities and ensuring that they are appropriately assessed. We hope that these studies of student learning, along with the formative tools and activities that supported them, provide convincing evidence for the importance of quality formative assessments in ethics education. When learning activities are properly aligned and assessed (i.e., including both formative and summative assessments), students can understand what they are supposed to learn and instructors can make informed adjustments to their instructional and curricular activities. While the importance of curricular alignment and assessment has been clearly demonstrated in the educational literature, for reasons discussed above, it appears to be a weakness in the present state of ethics education. We hope that the evidence; the example of learning, and practical tools presented here will help point the way toward building more effective designs.

**Acknowledgements:**

This work was supported by the National Science Foundation under grants IIS-0832843 and IIS-0832844. 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. We would like to thank Nicole Cooley for help with data coding; Ying Liu and Natalie Bolton for data analysis; and David Kijowski for his help in developing the case studies that were tested in these two studies.

**References:**

Angelo, T.A., & Cross, P. (1993). *Classroom Assessment Techniques* 2nd Ed. San Francisco, CA: Josey Bass.

Bebeau, M. J., & Thoma, S. (1999). Intermediate concepts and the connection to moral education. *Educational Psychology Review, (11)4, 343–360.*

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2008). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Black P. and Wiliam D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80(2), 139-148.*

Davis, M. & Feinerman, A. (2010) Assessing Graduate Student Progress in Engineering Ethics, *Science and Engineering Ethics,* doi 10.1007/s11948-010-9250-2

Davis, M., & Keefer, M. W. (2011) Getting Started: Helping a New Profession Develop an Ethics Program. *Science and Engineering Ethics.* doi: 10.1007/s11948-011-9279-x

DuBois, J. M, Schilling, D.A., Heitman , E., Steneck, N. H., & Kon, A. A., (2010). Instruction in the Responsible Conduct of Research: An Inventory of Programs and Materials within CTSAs. *Clinical and Translational Science (3)3, 109-111.*

Harris, C.E., Pritchard, M.S., & Rabins, M.J. (1995). *Engineering Ethics: Concepts and Cases* 4th Ed. Belmont CA: Wadsworth.

Kalichman, M. (2007). Responding to Challenges in Educating for the Responsible Conduct of Research. *Academic Medicine, (82)9, 870-875.*

Kalichman, M., & Plemmons, D. (2007). Reported Goals for Responsible Conduct of Research Courses. *Academic Medicine, (82)9, 846-852.*

Keefer, M. W., & Davis, M. (2012) Curricular Design and Assessment in Professional Ethics Education: Some Practical Advice. *Teaching Ethics. (13)1, 81-90.*

Keefer, M. W., & Wilson, S. (2011). Responding to Computational Modeling Cases: The importance of informal assessment. Paper presented at the 13th International Conference for the Society for Ethics Across the Curriculum. St. Louis, MO. November, 3-5.

Keefer, M.W., & Ashley, K.D. (2001) Case-based Approaches to Professional Ethics: A systematic comparison of students' and ethicists' moral reasoning. *The Journal of Moral Education. (30)4, 377-398.*

Kijowski, D. J. (2010). Responsible conduct of research with computational models and simulations. Illinois Digital Environment for Access to Learning and Scholarship (IDEALS). Retrieved May 25, 2011, from https://www.ideals.illinois.edu/handle/2142/18349

Kijowski, D.J., Dankowicz, H., & Loui, M.C., (2011) "Observations on the Responsible Development and Use of Computational Models and Simulations". Sci. and Eng. Ethics, To appear, available online, doi: 10.1007/s11948-011-9291-1.

Lehmann, L.S., Kasoff, W.S., Koch, P., & Federman, D.D. (2004). A Survey of Medical Ethics Education at U.S. and Canadian Medical Schools. *Academic Medicine, (79)7, 682-689.*

Loui, M. C. (2006) Assessment of an engineering ethics video: Incident at Morales," *Journal of Engineering Education, (95)1, 85–91.*

Resnick, L. B. (1987). Learning in school and out. *Educational Researcher, 16(9), 13-20.*

Stevens, D. D., and Levi, A. J. (2013). *Introduction to Rubics*. 2nd ed. Sterling, Va.: Stylus Publishing.

Wiggins, G., and McTighe, J. (2005). *Understanding by design* (Expanded 2nd Edition). Alexandria, VA: Association for Supervision and Curriculum Development.