## Notes in Brief

Implementation fidelity data indicate to what extent the delivered educational intervention (e.g., pedagogies, curricula) differs from the designed intervention (Gerstner & Finney, 2013; O'Donnell, 2008). Fidelity data help practitioners make more accurate inferences regarding program effectiveness (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001). However, implementation fidelity research is underused in higher education (Berman & McLaughlin, 1976; Dhillon, Darrow, & Meyers, 2015). Institutional and programmatic assessment cycles typically omit implementation fidelity processes. Moreover, there are too few didactic examples of how to engage in implementation fidelity (O'Donnell, 2008). Thus, we provide actionable steps for gathering implementation fidelity data. Practitioners who adopt these steps will be well-positioned to conduct fidelity research as part of assessment processes. They will also be able to draw more valid inferences from assessment data and make more informed decisions regarding interventions. Fidelity research can help higher education evolve from an assessment culture to a learning improvement culture.
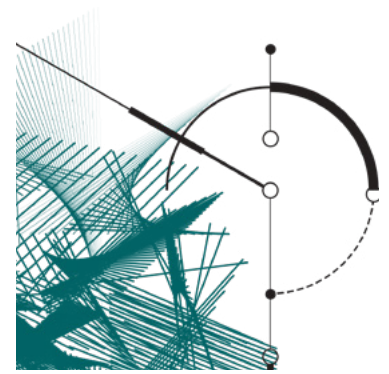
AUTHORS

Kristen L. Smith, Ph.D.
University of North Carolina-Greensboro

Sara J. Finney, Ph.D.
James Madison University

Keston H. Fulcher, Ph.D.
James Madison University

Actionable Steps for Engaging Assessment Practitioners and Faculty in Implementation Fidelity Research

I magine you are an assessment practitioner at a university. Several faculty members from a program on your campus decide to use their assessment results to improve students' learning. Specifically, previous assessment results indicated that students' ethical reasoning skills were weak. Faculty created a new educational intervention (i.e., new curricula and pedagogies) to improve students' ethical reasoning skills. The faculty attempted to implement the new educational intervention across several courses.

With your help, faculty collect student learning outcomes data via an ethical reasoning performance assessment instrument (e.g., a rubric). These data collection procedures involve standardized, rigorous, longitudinal methodology. The assessment instrument has been studied previously and found to be psychometrically sound (i.e., adequate reliability and validity evidence exist for scores). Therefore, student learning outcomes data are expected to be trustworthy and of high quality.

You analyze and present the student learning outcomes assessment data and results to faculty members. For some classes, students' ethical reasoning skills improved dramatically over the course of the semester (i.e., from before students experienced the new intervention to after they completed the new intervention). For other classes, students' ethical reasoning skills did not improve over time. The faculty ask: "Why? Please explain to us why students in some courses experienced great change in their ethical reasoning skills whereas students in other courses did not. What do the data indicate regarding *why* this occurred?"

To many assessment practitioners, these questions are all too familiar and the response "I don't know *why*" is often difficult to offer faculty who spent a great deal of time and

CORRESPONDENCE

Email
k_smith8@uncg.edu

energy collecting student learning outcomes data, with the goal of improving their program. Both faculty and assessment practitioners often leave these meetings feeling as though there is little that can actually be inferred or acted upon with respect to program improvement; student learning outcomes data (e.g., scores on performance assessment instruments, scores on multiple-choice tests) by themselves do not appear helpful.

Unfortunately, in these situations, this feeling is completely appropriate, as there is no data collected to help faculty understand *why* outcomes assessment results differed across classes. You have no information about what actually occurred in the classrooms when the new intervention was (supposed to be) implemented. Thus, the intervention that these faculty delivered can be thought of as a "black box." Inside this black box could be the intervention as it was designed or intended or an intervention that severely deviated from what was intended. Perhaps the intervention was delivered with higher quality in some classes compared to others, or students were more responsive in some classes but not in other classes. The black box obfuscates inferences about the designed program from the student learning outcomes assessment data. However, a specific line of inquiry exists to unlock this "black box" and facilitate more accurate inferences from student learning outcomes assessment data: implementation fidelity research.

> **More specifically, we describe the steps a group of faculty took when engaging in implementation fidelity research. By detailing these steps we aim to promote more implementation fidelity research within higher-education contexts.**

Implementation fidelity has been defined as "the degree to which a program model [educational intervention] is instituted as intended" (Dhillon, Darrow, & Meyers, 2015, p. 9). Other names for implementation fidelity include enacted curriculum, program integrity, treatment integrity, and clinical effectiveness (Dhillon, Darrow, & Meyers, 2015; Mellard, 2010). Implementation fidelity data indicate to what extent the *delivered* educational intervention (e.g., pedagogies, curricula) differs from the *designed* or *planned* educational intervention (Gerstner & Finney, 2013; O'Donnell, 2008). The five components of implementation fidelity data include:

- specific features and components of the intervention (i.e., program differentiation),
- whether each feature or component was actually implemented (i.e., adherence),
- quality with which features and components of the intervention were implemented,
- perceived student responsiveness during implementation, and
- duration of implementation.

O'Donnell (2008) and Gerstner and Finney (2013) provide more detailed definitions of implementation fidelity. Implementation fidelity data is often collected via class observations using a *fidelity checklist*, which is a behaviorally based data collection instrument, as described in greater detail by Swain, Finney, and Gerstner (2013). Although the aforementioned articles offer definitions, describe data collection tools, and emphasize the need to collect implementation data, few resources guide practitioners through the entire process of implementation assessment via an applied example.

Purpose

This article's purpose is to provide a didactic example guiding practitioners and faculty through the process of gathering implementation fidelity data. More specifically, we describe the steps a group of faculty took when engaging in implementation fidelity research. By detailing these steps we aim to promote more implementation fidelity research within higher-education contexts.

Implementation fidelity data can be gathered for virtually any educational content area at an institution of any size (Durlak & DuPre, 2008). We describe how implementation fidelity data were gathered for a campus-wide ethical reasoning initiative at James Madison University (JMU), *The Madison Collaborative: Ethical Reasoning in Action*. Because the focus of this article is implementation fidelity research, not ethical reasoning, we do not elaborate on how ethical reasoning was defined or assessed by the Madison Collaborative. Nevertheless, for readers interested in these details, we refer them to Ames et al. (2016) and Sanchez, Fulcher, Smith, Ames, and Hawk (2017).

The implementation fidelity practices described in this article contributed to large-magnitude student learning improvement across multiple courses, disciplines, and student developmental levels at JMU. That is, gains in student learning were greater in courses where the educational intervention was implemented with high fidelity compared to courses where the intervention was implemented with lower fidelity. In a forthcoming, separate article, we describe how fidelity data were integrated with student learning outcomes assessment data to facilitate and demonstrate learning improvement. Our goal is to detail the implementation fidelity process itself, showcasing how fidelity data—on their own—can be powerful for understanding the educational intervention students receive and necessary if learning improvement is the goal (Finney & Smith, 2016).

Importance of Implementation Fidelity Research

Prior to detailing the steps of gathering implementation fidelity data employed by the faculty on our campus we explain the importance of fidelity data. In brief, fidelity data are crucial for modifying educational interventions and demonstrating learning improvement (Fisher, Smith, Finney, & Pinder, 2014). Implementation fidelity data provide important information that can enhance the accuracy of the inferences made from student learning outcomes assessment data (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001). Fidelity data also allow faculty to more systematically understand the educational intervention that their students actually received. Moreover, Durlak and DuPre (2008) concluded that high fidelity of implementation contributes to the success of educational programming [interventions].

More specifically, when student learning outcomes assessment data or results are unfavorable faculty are left wondering why. With implementation fidelity data practitioners and faculty are equipped to explain "why" and make informed changes to the educational intervention. That is, perhaps student learning outcomes assessment data were unfavorable because an intervention feature was not actually implemented or an intervention feature was delivered with low quality (Dhillon, Darrow, & Meyers, 2015). If so, implementation fidelity data can help faculty backward design courses (Fink, 2003), enhancing alignment between assessment, pedagogy, curriculum, and student learning. Alternatively, when student learning outcomes assessment data are favorable (e.g., students' scores improve), implementation fidelity data can "provide a roadmap for replication" and help identify "critical ingredients of program success" (Bond, Evans, Salyers, Williams, & Kim, 2000, p. 79). Understanding the effectiveness of intervention features allows faculty to be more pedagogically efficient and intentional. They can avoid "wasting" time on features of an intervention that have been shown to be ineffective for student learning improvement.

In contrast, without implementation fidelity data, it is difficult to determine whether unfavorable assessment results are due to a poorly designed intervention or incomplete/inadequate delivery of the designed intervention (Dhillon, Darrow, & Meyers, 2015). Lack of fidelity data can lead faculty to make one of two costly errors:

- abandoning effective interventions (that perhaps were not implemented with high fidelity), or
- continuing to implement ineffective interventions (Gerstner & Finney, 2013).

In addition to these errors, lack of fidelity data can contribute to invalid inferences. For example, if student learning outcomes assessment data indicate that students' ethical reasoning skills improved from the beginning to the end of the semester faculty may (incorrectly) conclude that their new educational intervention was effective. In reality, the faculty did not implement the new intervention with high fidelity, and thus the new intervention cannot be credited with contributing to improvements in students' knowledge or abilities. Implementation fidelity provides important information that enhances the accuracy of the inferences made from student learning outcomes assessment data (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001; Fisher, Smith, Finney, & Pinder, 2014). As described by Durlak and DuPre, "without data on implementation, research cannot document precisely what program [educational intervention] was conducted, or how [student learning] outcome data should be interpreted" (2008, p. 340).

Fidelity data also allow faculty to more systematically understand the educational intervention that their students actually received.

For higher education practitioners who require external funding to support assessment efforts implementation fidelity data is becoming increasingly important. The U.S. Department of Education requires grant recipients to measure and report implementation fidelity to gauge educational program impact (Goodson, Price, & Darrow, 2015). In addition, public and private organizations are funding research to examine fidelity in educational contexts, develop best practices for fidelity research, and refine how fidelity is measured (Dhillon, Darrow, & Meyers, 2015; Hulleman & Cordray, 2009). Medical researchers have been measuring implementation fidelity for years (Bond et al., 2000; Rogers, Eveland, & Klepper, 1977).

Unfortunately, implementation fidelity continues to be underused in educational research—especially higher education (Berman & McLaughlin, 1976; Dhillon, Darrow, & Meyers, 2015). Although fidelity data are imperative for assessment best practices, the collection, analysis, and integration of implementation fidelity data are completely absent from most institutional and programmatic assessment cycles. Instead, assessment practitioners incorrectly assume that the "delivered" or "implemented" intervention is the same as the "designed" or "intended" intervention. Moreover, assessment practitioners mistakenly infer that student learning outcomes were achieved as a result of the "intended" educational intervention, not what actually occurred in the classrooms (i.e., the "delivered" intervention). This misconception is not surprising.

> **Although fidelity data are imperative for assessment best practices, the collection, analysis, and integration of implementation fidelity data are completely absent from most institutional and programmatic assessment cycles.**

Method

O'Donnell (2008) notes a lack of literature guiding practitioners through the implementation fidelity process. How can assessment practitioners engage in implementation fidelity research more frequently and effectively without instructive examples of how to do so? In response to that question we detail the following steps our faculty followed to collect implementation fidelity data for an ethical reasoning educational intervention. We also highlight how fidelity data on their own (i.e., before they are integrated with student learning outcomes assessment data) can help faculty understand which features of the intervention students received (Finney & Smith, 2016).

Step 1: Allocate Adequate Time, Space, and Expertise

Implementation fidelity research requires several inputs from assessment practitioners and faculty (e.g., a targeted student learning outcome, an educational intervention, a fidelity checklist, an assessment instrument or tool). Adequate time must be set aside for creation of these materials. On our campus, six faculty from diverse disciplines and backgrounds participated in a week-long training institute related to implementation fidelity and student learning improvement. The institute took place during the summer. As detailed in Appendix A, activities for the institute included:

> **On our campus, six faculty from diverse disciplines and backgrounds participated in a week-long training institute related to implementation fidelity and student learning improvement.**

- helping faculty members understand implementation fidelity research processes,
- helping faculty understand the assessment instrument (i.e., the rubric) used to evaluate students' ethical reasoning skills and how that was related to fidelity research,
- providing examples of fidelity research studies,
- reviewing implementation fidelity checklists and fidelity data collection processes,
- helping faculty articulate their program theory,
- allowing faculty to draw from their own experiences and learning activities to create a new learning intervention, and so forth.

During the institute, assessment practitioners used group discussions, "think. pair. share." and other activities to engage faculty. Readers are encouraged to review Appendix A for a more detailed explanation of the institute's structure and specific content. The activities included in Appendix A can be used as a template to help practitioners provide adequate time, space, and expertise to faculty as they begin engaging in fidelity research.

Faculty were compensated for their time during the institute; however, they were also intrinsically motivated to participate (e.g., faculty indicated interest in participating in the training institute before knowing a stipend would be provided). Assessment experts used group activities, peer-to-peer feedback, and other tools to promote a collaborative and safe environment for faculty. Figure 1 provides a snapshot of the processes used during the training institute to help faculty engage in implementation fidelity research.
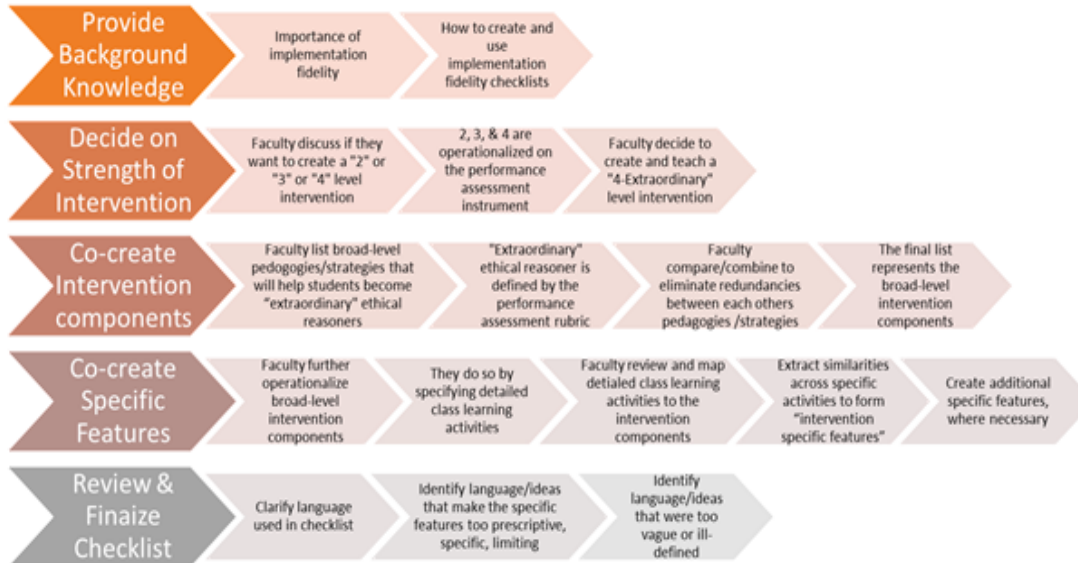

Figure 1. Visualization of Process Used During Training Institute to Help Faculty Create an Ethical Reasoning Intervention and Accompanying Fidelity Checklist

We created the following learning outcomes for the training institute. As a result of participating in the 2016 Implementation Fidelity Research Training Institute faculty will:

- Explain how assessment practice and teaching and learning are connected or related.

- Identify the five components of implementation fidelity.

- Explain the steps of collecting implementation fidelity data.

- Articulate why implementation fidelity data is important for demonstrating student learning improvement.

- Discuss and agree upon the specific features of an effective ethical intervention aligned with one of the James Madison University's Madison Collaborative ethical reasoning student learning outcomes.

- Design an ethical reasoning intervention based on the agreed upon features that aligns with the targeted Madison Collaborative ethical reasoning student learning outcome and that can be applied in various classes.

- Create a general implementation fidelity checklist aligned with the ethical reasoning intervention and the targeted university ethical reasoning student learning outcome.

These faculty learning outcomes were used to prepare and deliver institute activities and content. As shown in Appendix A, each institute activity was mapped back to at least one of the faculty learning outcomes.

In addition to the learning outcomes, the institute had two main deliverables. First, faculty were charged with detailing the specific components and features of an ethical reasoning educational intervention that they all agreed to implement within their respective classes. Second, faculty were asked to create an accompanying implementation fidelity checklist. This

checklist would later be used to capture the extent to which the designed intervention was actually delivered to students.

Faculty participants needed appropriate assessment and subject matter expertise to accomplish the learning objectives, create an educational intervention, and build a fidelity checklist. A team of two assessment practitioners, along with an ethical reasoning subject matter expert, worked closely with the faculty. At least one assessment and/or subject matter expert guided faculty through various presentations and working sessions each day of the institute[1]. The assessment practitioners were affiliated with our campus Center for Assessment and Research Studies. Our campus also has a separate Center for Faculty Innovation who provided the physical space where the summer training institute took place. The assessment practitioners worked closely with the Center for Faculty Innovation on other projects, and thus had received some cross-training in faculty development best practices.

Step 2: Facilitate Faculty Understanding of Targeted Student Learning Outcome and Assessment Instrument(s)

Before faculty could create an educational intervention, they needed to understand the learning outcome targeted for improvement. That is, effective educational interventions are intentionally created to impact particular skills or abilities. An intervention built to impact one outcome may not be effective at impacting another outcome. These ideas were discussed with the faculty.

The student learning outcome that faculty targeted for improvement concerned students' abilities to *apply their ethical reasoning skills to their personal, professional, and civic lives*. Thus, faculty needed to discuss and process what it means for students to apply their ethical reasoning skills. To facilitate this processing, an assessment expert began by explaining the importance of this outcome at the program- and university-levels.

Faculty then familiarized themselves with the rubric used to assess students' achievement of the targeted learning outcome. The rubric was a locally developed instrument designed to measure students' application of ethical reasoning skills (see Appendix B). Researchers had previously studied this rubric. As a result of those studies, reliability and validity evidence for rubric scores was provided (Smith, Pyburn, & Ames, 2016; Sanchez, Fulcher, Smith, Ames, & Hawk, 2017).

An assessment expert provided copies of the ethical reasoning performance assessment rubric and reviewed all rubric criteria with the faculty. Multiple faculty members previously used the assessment rubric to rate students' ethical reasoning essays. These faculty members were asked to share their experiences using the rubric, their insights about what the rubric was measuring, and their interpretations of the rubric's criteria. The assessment expert also shared previous years' assessment results to help faculty understand the extent to which students were achieving the targeted learning outcome. Using the assessment rubric to help define and clarify the outcome promoted alignment between the assessment instrument (i.e., the rubric) and the educational intervention. The assessment rubric also provided a common language and crucial reference point for faculty who were approaching ethics education from diverse backgrounds and experiences.

Step 3: Facilitate Faculty Understanding of Implementation Fidelity

At this point, faculty understood the targeted student learning outcome they wanted to improve (i.e., ethical reasoning) and had studied the assessment instrument used to measure those skills (i.e., the ethical reasoning rubric provided in Appendix B). Now they were ready to study best practices in implementation fidelity. First, we explained the importance of implementation fidelity research, as well as identified and extensively discussed the five

> **At this point, faculty understood the targeted student learning outcome they wanted to improve (i.e., ethical reasoning) and had studied the assessment instrument used to measure those skills (i.e., the ethical reasoning rubric provided in Appendix B). Now they were ready to study best practices in implementation fidelity.**

[1]Note, this type of faculty development does not necessarily have to occur during a week-long institute. However, providing adequate time and space for faculty education, discussion, creation, etc. is imperative to engaging in implementation fidelity research. Moreover, consider working with external consultants if you do not have access to assessment practitioners or a subject matter expert on your campus. It is important to provide faculty with appropriate expertise as they create fidelity checklists and engage in implementation fidelity research.

components of implementation fidelity (O'Donnell, 2008; Gerstner & Finney, 2013). Faculty reviewed at least three different examples of fidelity checklists and asked questions to clarify their understanding of the five components of implementation fidelity research. As discussed previously, these five components included specific features and components of the intervention, whether each feature or component was actually implemented, the quality with which features and components of the intervention were implemented, perceived student responsiveness during implementation, and duration of implementation. We introduced faculty to five different implementation fidelity data collection methodologies and discussed how implementation fidelity data on their own (i.e., before they are integrated with student learning outcomes assessment data) are extremely useful for articulating the educational intervention students actually receive (Finney & Smith, 2016). We then provided numerous examples of how implementation fidelity data can be coupled or integrated with student learning outcomes assessment data to make more accurate inferences about student learning. This information was conveyed primarily through presentations and small group discussions.

Several faculty participants had no prior knowledge of implementation fidelity research. Thus, it was important to spend adequate time familiarizing them with these processes. Moreover, we reiterated that implementation fidelity data would not be used in an evaluative or punitive way (e.g., to evaluate their teaching prowess, make decisions about tenure). This frank discussion helped alleviate faculty concerns about potential uses of fidelity data while continuing to support an innocuous environment.

Step 4: Guide Faculty Through Creation of Educational Intervention and Fidelity Checklist

After faculty participants understood implementation fidelity research best practices the assessment and subject matter experts guided faculty as they built a new ethical reasoning educational intervention. Recall, the newly created intervention was constructed under the guiding framework of the ethical reasoning performance assessment rubric (see Appendix B). The assessment expert asked faculty what strength of educational intervention they wanted to create, in reference to the ethical reasoning performance assessment rubric. For instance, she asked faculty if they wanted to create an intervention that would facilitate their students being able to demonstrate "3-Excellent" ethical reasoning skills or "4-Extraordinary" ethical reasoning skills. Note, this questioning was intentionally and explicitly linked to the criteria and elements detailed on the assessment rubric. The intent was to facilitate alignment between the educational intervention, the targeted learning outcome, and the assessment instrument (i.e., the ethical reasoning rubric provided in Appendix B).

Faculty decided that they wanted to create an intervention that would help students demonstrate "4-Extraordinary" ethical reasoning skills. Thus, while faculty were creating their ethical reasoning intervention they had clear and common criteria detailed in the assessment rubric to guide them (see Appendix B). They understood they now needed to create an intervention that supported students becoming "4-Extraordinary" ethical reasoners as defined by the characteristics and skills noted in the rubric. They returned to the rubric continually as they built and rebuilt the new educational intervention.

At this stage, it was important to help faculty think through their program theory. Program theory provides a model of how a given educational intervention is expected to work (Rogers, Petrosino, Huebner, & Hacsi, 2000). Expanding on Bickman's (1987) conceptualizations, faculty members should create and articulate a program theory which details the specific aspects of their educational intervention and how that intervention is supposed to work—in theory—to enhance student learning, help students acquire a certain skillset, and more. The program theory is in reference to specific outcomes (i.e., criteria). The purpose of conducting outcomes assessment is to understand if the educational intervention— which is operationalizing a clearly articulated theory of how students should acquire certain knowledge and skills—is effective. The program theory explains *why* and/or *how* certain intervention specific features should result in students achieving certain learning outcomes.

To help articulate their program theory faculty generated a list of intervention components or "broadly-stated activities, pedagogies or approaches" that they could integrate into their classes that *should* help students become level "4-Extraordinary" ethical reasoners (See "Co-Create Intervention Components" in Figure 1). The assessment expert asked faculty

After faculty participants understood implementation fidelity research best practices the assessment and subject matter experts guided faculty as they built a new ethical reasoning educational intervention.

to provide rationale for *why* these components *should*—in theory—help students improve their ethical reasoning skills. Faculty referenced literature from cognitive psychology to help provide such rationale (e.g., Halpern & Hakel, 2003).

Faculty then participated in a series of "think. pair. share." exercises to co-create the components of the intervention. Intervention components are broadly specified activities, pedagogies, or curriculum, and the specific features operationalize or detail the activities under each component. Faculty compared and contrasted each other's intervention components to eliminate redundancies where appropriate.

To then operationalize these broad intervention components faculty began by sharing specific activities, assignments, demonstrations, case studies, or other learning opportunities they implemented in their classes in the past, or planned to implement in the future, to help students achieve "4-Extraordinary" ethical reasoning skills. Each faculty shared these activities with one partner, refined them, and then presented to the larger group. Then we helped faculty categorize, or map, all of the specific activities (i.e., intervention-specific features) to the intervention components that they previously articulated. The intervention-specific features were edited to be more generalizable, such that each specific feature of the intervention would be general enough to be applied across the different courses and disciplines of each faculty member.

For example, "case studies" was one of the intervention components that faculty thought would be important for helping students become "4-Extraordinary" ethical reasoners. Several faculty shared specific assignments and/or activities from their classes that would be aligned with the case studies component. As a group, faculty took these course- and discipline-specific assignments and/or activities and pulled out any underlying commonalities or similarities. These common threads became the specific features on the intervention implementation fidelity checklist (see Appendix C).

We encourage readers to review the fidelity checklist in Appendix C to understand the specific features of the intervention that faculty co-created during the summer training institute. The checklist is a vital tool for fidelity research because it details the specific features of the educational intervention, and aligns those features to student learning objectives (Swain, Finney, & Gerstner, 2013). Readers can use the checklist provided in Appendix C as a template for helping faculty articulate their own program theory and build well-aligned educational interventions. Furthermore, the checklist can serve as a template for numerous constructs or content areas of interest other than ethical reasoning.

Once the intervention components and specific features were articulated faculty critically reviewed them. They clarified language in the intervention components and specific features, identifying any instances where language/ideas were too prescriptive, specific, or limiting, as well as instances where language/ideas could be further detailed. The goal was to create an intervention that, if effective, could be easily understood and implemented by other faculty, in a variety of classes. The specific features and components of the intervention were finalized and used to create an implementation fidelity checklist, as shown in Appendix C. The fidelity checklist was general enough to be used across a wide variety of classes to collect fidelity data related to students' abilities to apply their ethical reasoning skills (i.e., the student learning outcome targeted for improvement).

Step 5: Co-create a Fidelity Data Collection Plan with Faculty

Fidelity researchers used the checklist throughout the fall 2016 semester to collect fidelity data from all six faculty participants' classes. The implementation fidelity checklist was converted into an excel worksheet, facilitating electronic gathering and storage of fidelity data. Collecting and storing fidelity data in electronic format, as opposed to paper-pencil, simplified the process of adjudicating and integrating the fidelity data with the student learning outcomes assessment data.

Note, in accordance with institutional IRB procedures, faculty participants signed an informed consent form granting consent for fidelity researchers to observe their classrooms and collect fidelity data using the fidelity checklist. During these class observations researchers

The goal was to create an intervention that, if effective, could be easily understood and implemented by other faculty, in a variety of classes. The specific features and components of the intervention were finalized and used to create an implementation fidelity checklist.

applied the checklist to at least six specified class sessions and/or specified class assignments throughout the semester. Fidelity researchers discussed and adjudicated fidelity data (e.g., came to agreement, averaged scores) to ensure that one researcher did not overlook any specific features that were implemented, or that one researcher did not rate quality of implementation too low, etc.

In addition to fidelity researchers observing class sessions to gather fidelity data, each faculty member filled in the checklist for him or herself as a self-report indication of fidelity (i.e., a "self-audit") for at least three class sessions throughout the semester. Researchers asked faculty to complete self-audits for several reasons. First, self-audits provided additional implementation fidelity data points (i.e., in addition to those collected by the fidelity researchers who observed class sessions). Additional data points promoted greater accuracy of fidelity data. Data from faculty self-audits were used in data adjudication processes described previously. Second, self-audits allowed faculty to further engage in the fidelity research process by collecting fidelity data on their own class sessions. Faculty were able to contribute their own fidelity data points to the larger pool of data points being collected by the fidelity researchers. Lastly, asking faculty to complete self-audits encouraged them to review the checklist and remain familiar with the specific intervention features they had decided to implement.

Faculty were permitted to complete the self-audit checklists during two different occasions, depending on what was most feasible for them. For example, instructors could fill out the checklist for themselves on occasions when the fidelity researchers were not able to attend the class to collect fidelity data. Thus, faculty were able to capture fidelity data points that would have otherwise been missed due to fidelity researchers not being able to observe the class session. Alternatively, faculty could fill out the checklist for themselves during class sessions where fidelity researchers were able to observe and collect fidelity data. In this instance, the faculty self-audit fidelity data provided additional data points that were adjudicated with fidelity researchers' data to enhance accuracy. Additionally, some faculty filled out the checklist via paper-pencil, whereas others filled it out electronically using an excel worksheet depending on their preference. An assessment specialist converted all paper-pencil data to electronic form.

Having faculty complete self-audits, in addition to fidelity researchers collecting fidelity data, is considered best practice (Gerstner & Finney, 2013). First, self-audit practices provide additional data points that enrich interpretation of results. When fidelity data from faculty self-audits are consistent with fidelity data collected via fidelity researchers' observations there is initial evidence of data trustworthiness. Second, self-auditing can protect against program drift by explicitly reminding faculty of the specific features they intended to include in the intervention (Gerstner & Finney, 2013). Third, engaging in self-auditing promoted faculty buy-in for implementation fidelity processes. Faculty also demonstrated greater interest in student learning outcomes assessment results, given their personal time spent collecting fidelity data via self-audits.

Step 6: Share and Discuss Fidelity Data with Faculty

To promote transparency all fidelity data were shared with faculty for review. After reviewing the fidelity data for each class faculty provided feedback to ensure data accuracy. For example, faculty made note of any specific features that were implemented that the fidelity researcher might have missed and commented on whether perceived student responsiveness ratings seemed accurate. The fidelity data review processes were important given fidelity data are observational. Even the best-trained fidelity researchers occasionally miss an intervention feature being implemented, misinterpret student responsiveness during class, etc.

After faculty reviewed fidelity data for accuracy a fidelity researcher synthesized and summarized all fidelity data for each faculty member across the entire semester. The fidelity researcher shared these summaries (e.g., graphs and tables of fidelity data) with the faculty, individually and as a larger group. Thus, faculty could easily evaluate the degree to which the designed educational intervention was actually implemented across their various classes. Moreover, faculty could understand what intervention features their students actually

Collecting and storing fidelity data in electronic format, as opposed to paper-pencil, simplified the process of adjudicating and integrating the fidelity data with the student learning outcomes assessment data.

received compared to the features that students from other classes received. Given the faculty were involved in articulating the intervention features, creating the checklist, collecting the fidelity data, etc., they understood and appreciated the quality of these fidelity data and their utility for drawing accurate conclusions about the efficacy of the new learning intervention.

The fidelity data suggested that the intervention was implemented with varying degrees of fidelity across different faculty members' classes. For example, five of the six faculty were able to implement (i.e., "adhered to") most of the specific features on the checklist (see Appendix C). However, there was some variability in student perceived responsiveness and quality of implementation. One faculty member's class had exceptionally high levels of student perceived responsiveness compared to other faculty. This same faculty member implemented the intervention with high fidelity and used the greatest variety of activities, exercises, and so forth to implement the intervention-specific features. Duration of implementation and adherence differed notably among faculty members. Some faculty members implemented the specific features with greater frequency compared to other faculty. For instance, one faculty member implemented the specific features with much greater frequency than the other faculty, but their students were perceived to be less responsive during implementation. Fidelity data also indicated that certain intervention-specific features were very rarely (or in some classes never) implemented. Perhaps faculty need further development or training to effectively implement these features, or perhaps these features are not salient to an effective educational intervention.

Overall, fidelity data demonstrated that the new educational intervention could be implemented with moderate to high fidelity across a variety of disciplines, course-types (e.g., large v. small, lecture v. community service learning, etc.), and contexts. Fidelity data on their own (i.e., before they were integrated with student learning outcomes assessment data) were powerful for understanding the educational intervention that students received (Finney & Smith, 2016). Nevertheless, the next step was to integrate fidelity data with student learning outcomes assessment data (i.e., students' scores on the performance assessment rubric provided in Appendix B) to evaluate student learning associated with the new intervention (Fisher, Smith, Finney, & Pinder, 2014). Through the process of integration, fidelity data allowed faculty to examine why students' rubric scores improved differentially across various classes over the course of the semester. This speaks directly to the hypothetical situation described at the opening of this article.

Fidelity data illuminated the black box. Faculty and researchers were able to link differential improvements in students' learning back to what students actually experienced in the classroom. Once fidelity data were integrated with the student learning outcomes assessment data (i.e., students' pre- and post-test scores on the performance assessment rubric) faculty were able to understand how certain features of the educational intervention may have contributed to students' learning improvements. In a forthcoming article we explain in greater detail how fidelity data were integrated with student learning outcomes assessment data.

Once fidelity data were integrated with the student learning outcomes assessment data (i.e., students' pre- and post-test scores on the performance assessment rubric) faculty were able to understand how certain features of the educational intervention may have contributed to students' learning improvements.

Conclusion

When differential learning gains were observed across classes, instead of asking "*why*," assessment practitioners and faculty turned to implementation fidelity data to explain the results. This was empowering, as it allowed faculty to identify which intervention-specific features were implemented with high fidelity, how students' perceived responsiveness contributed to learning gains, etc. The implementation fidelity practices described in this article also contributed to large-magnitude student learning improvement across multiple courses, disciplines, and contexts. That is, students' ethical reasoning rubric scores (see Appendix B), on average, improved two standard deviations (*Cohen's d = 2*) from the beginning to the end of the semester. According to Cohen (1988) the threshold for a large effect is 0.8. In this context, the magnitude of improvement in students' ethical reasoning abilities was exceptionally large. The new ethical reasoning intervention—articulated and studied via the implementation fidelity steps described previously—was found to be effective.

Yet, the overall effect does not tell the whole story. In a subsequent article we will explicitly describe how we integrated implementation fidelity data with student learning outcomes assessment data. This integration process allowed us to make fine-tuned inferences about which intervention features worked and which did not. Further, we were able to make recommendations regarding how the intervention could be strengthened and how faculty could deliver it more effectively in the future.

Granted, it is not enough to explain the steps needed to begin engaging in implementation fidelity research. Assessment practitioners and faculty need further guidance on how to:

- analyze fidelity data,
- integrate fidelity data with student learning outcomes assessment data, and
- present integrated data to faculty in a way that is meaningful and actionable (Coburn, Hill, & Spillane, 2016; O'Donnell, 2008).

This was beyond the scope of the current article but will be provided in forthcoming work. We encourage assessment practitioners to engage with implementation fidelity to help others understand how it can facilitate learning improvement. We know that assessment—as currently practiced—has produced few examples of learning improvement (Banta & Blaich, 2011; Blaich & Wise, 2011; Fulcher, Good, Coleman, & Smith, 2014). Perhaps we should stop griping about the under-use of learning outcomes assessment results and start investigating the efficacy of educational interventions via implementation fidelity research.

Implementation fidelity cracks open the black box of higher-education curriculum and pedagogy. With such a link, assessment can help close the learning improvement loop. Without the connection to curriculum and pedagogy—provided by fidelity data—assessment merely perseverates a data collection loop.

We encourage assessment practitioners to engage with implementation fidelity to help others understand how it can facilitate learning improvement.

References

Ames, A. J., Smith, K. L., Sanchez, E. R. H., Pyle, L. K., Ball, T. C., & Hawk, W. J. (2016). Impact and persistence of ethical reasoning education on student learning: Results from a module-based ethical reasoning educational program. *International Journal of Ethics Education*. doi:10.1007/s40889-016-0031-x.

Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning, 43*(1), 22–27.

Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovation. *The Educational Forum, 40*(3), 345–370.

Bickman, L. (1987). The functions of program theory. *New directions for program evaluation, 33*, 5–18.

Blaich, C. F., & Wise, K. S. (2011, January). *From gathering to using assessment results: Lessons from the Wabash National Study* (NILOA Occasional Paper No. 8). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H. W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research, 2*(2), 75–87.

Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The common core state standards and implementation research. *Educational Researcher, 45*(4), 243–251.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Lawrence Erlbaum Associates.

Dhillon, S., Darrow, C., & Meyers, C. V. (2015). Introduction to implementation fidelity. In C. V. Meyers and W. C. Brandt (Eds.), *Implementation fidelity in education research* (pp. 8–22). New York, NY: Routledge.

Dumas, J., Lynch, A., Laughlin, J., Smith, E., & Prinz, R. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the early alliance prevention trial. *American Journal of Preventative Medicine, 20*(1S), 38–47.

Durlak, J. A., & DuPre E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.

Finney, S. J., & Smith, K. L. (2016). Ignorance is not bliss: Implementation fidelity and learning improvement. *National Institute for Learning Outcomes Assessment: Guest Viewpoints*. Retrieved from: https://illinois.edu/blog/view/915/309716

Fisher, R., Smith, K. L., Finney, S. J., & Pinder, K. (2014). The importance of implementation fidelity data for evaluating program effectiveness. *About Campus, 19*, 28–32.

Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014, December). A simple model for learning improvement: Weigh pig, feed pig, weigh pig. (NILOA Occasional Paper No. 23). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Gerstner, J. J., & Finney, S. J. (2013). Measuring the implementation fidelity of student affairs programs: A critical component of the outcomes assessment cycle. *Research and Practice in Assessment, 8*, 15–28.

Goodson, B., Price, C., & Darrow, C. (2015). Measuring fidelity. In C. V. Meyers and W. C. Brandt (Eds.), *Implementation fidelity in education research* (pp. 176–193). New York, NY: Routledge.

Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning. *Change Magazine, 35*(4). 36–41.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88–110.

Mellard, D. (2010). *Fidelity of implementation within a response to intervention framework*. Retrieved from: http://ped.state.nm.us/ped/RtIdocs/Fidelity%20of%20Implementation%20guidev5.pdf

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84.

Rogers, E. M., Eveland, J. D., & Klepper, C. (1977). *The Innovation Process in Organizations*. Department of Journalism, University of Michigan, Ann Arbor, MI.

Rogers, P. J., Petrosino, A., Huebner, T. A., & Hacsi, T. A. (2000). Program theory evaluation: Practice, promise, and problems. *New Directions for Evaluation, 87*, 5–13.

Sanchez, E. R. H., Fulcher, K. H., Smith, K. L., Ames, A. J., & Hawk, W. J. (2017, March-April). Defining, teaching, and assessing ethical reasoning in action. *Change: The Magazine of Higher Learning, 49*(2), 30–36.

Smith, K. L., & Pyburn, L., & Ames, A. J. (2016). *Madison Collaborative Annual Assessment Report #3*, James Madison University, Harrisonburg, VA.

Swain, M. S., Finney, S. J., & Gerstner, J. J. (2013). A practical approach to assessing implementation fidelity. *Assessment Update, 25*(1), 5–7, 13.

Appendix A

Faculty Summer Training Institute
Schedule At-a-Glance

	Activities/Curriculum	**Faculty Learning Outcomes**

Day 1 → **Implementation Fidelity Basics** ∙∙∙∙∙∙∙∙∙∙∙∙∙ **Understanding and Applying Implementation Fidelity**	• Brief introduction to the research project: explain why we are all here; the need for this work; review faculty signed MOUs and faculty responsibilities/roles in the project
	 • Brief introduction to assessment cycle
	 • Introduce implementation fidelity through examples from James Madison University's campus and introduce very general idea of backward design
	 • Discuss the five components of Implementation Fidelity
	   o Think. Pair. Share- Work with partner to fill in a blank implementation fidelity checklist for one intervention that you do in your class (can pick any intervention/activity/assignment, etc.)
	   o What was the hardest part about creating the checklist? What components require further clarification?
	   o Explain how implementation fidelity information can be useful pedagogically and useful for demonstrating learning improvement
	 • Describe the typical Implementation Fidelity data collection process
	   o The James Madison University Orientation Program
	   o The James Madison University LID CIS project
	 • Group discussion about the implementation fidelity matrix of possible inferences (Gerstner & Finney, 2013)
	   o Work through four (hypothetical) examples set in an academic contexts using the fidelity matrix (Gerstner & Finney, 2013) to convey the importance of fidelity data when making inferences based on outcomes assessment data
	 Day 1 Wrap Up:
	 Call back to why we are here: to apply implementation fidelity principles to Ethical Reasoning Instruction; to give faculty members development opportunities and skills that they can use beyond this research project. Tomorrow we will review the James Madison University ethical reasoning objective and discuss ethical reasoning educational interventions
		• Describe the steps of the assessment cycle
		 • Explain how assessment practice and teaching and learning are connected or related
		 • Identify the five components of implementation fidelity
		 • Explain the steps or process of collecting implementation fidelity data
		 • Articulate why implementation fidelity data is important for demonstrating student learning improvement
		 • Create a "general" implementation fidelity checklist aligned with the ER intervention and JMU's ethical reasoning student learning outcome

Days 2,3, & 4 → **Application of Implementation Fidelity to Ethical Reasoning (ER) Education** ∙∙∙∙∙∙∙∙∙∙∙∙∙ **Creating an ethical reasoning Intervention &** **accompanying fidelity checklist mapped to James Madison University's ethical reasoning student learning outcome**	• Brief review of the "program differentiation" component of implementation fidelity
	 • Brief review of the Ethical Reasoning 8 Key Questions
	 • Review the university ethical reasoning student learning outcome targeted for improvement & the pre-existing institution-wide interventions that are mapped to each
	 • Think. Create. Pair. Share: Individually, articulate the key features of what you believe would be a "highly effective" ethical reasoning intervention aligned with targeted university ethical reasoning student learning outcome that you could do in your classroom. Discuss in small groups and as larger group
	   o In order for students to be able to do the targeted learning outcome, what do we need to have them practice in our classrooms? What general things or "key features" must students do in order to achieve the targeted ethical reasoning learning outcome? How can these be generalized across disciplines? How can I teach students these things or integrate these "key features" into my course?
	 • Integrate these key features into a clear, agreed upon list of key intervention features
	   o General "Key features" must be agreed upon by all faculty participants
	 • Provide "blank" fidelity checklist and have faculty fill in with agreed upon key features
	   o This will be the final checklist used for data collection
		• Discuss and agree upon key components or features of an effective ethical reasoning intervention aligned with JMU's ethical reasoning student learning outcome
		 • Based on those agreed upon components, design an ethical reasoning intervention aligned with JMU's ethical reasoning student learning outcome that can be applied in various classes
		 • Create a "general" implementation fidelity checklist aligned with the ethical reasoning intervention and JMU's ethical reasoning student learning outcome

Day 5 → ∙∙∙∙∙∙∙∙∙∙∙∙∙ **Finalizing ethical reasoning intervention, checklist, & creating fidelity data collection plan**	• Faculty complete filling in fidelity checklist with agreed upon key features • Create implementation fidelity data collection procedures for Fall 2016 • Create schedule for when researchers will observe classes to collect implementation fidelity data • Discuss expectations for faculty "self-audit" using the fidelity checklist	

Appendix B

Performance Assessment Rubric Used to Rate Student Ethical Reasoning Essays

Insufficient 0	Marginal 1	Good 2	Excellent 3	Extraordinary 4
No reference to decision option(s).	Implicit reference to decision options AND/OR little context given regarding decision option(s).	Explicit but unorganized reference to decision option(s) and context.	Clear description of decision option(s) and context.	Meets criteria for *Excellent* AND… • Context treated with nuance. • Builds tension with organization and word choice.
Reference to zero or only one key question.	Vague references to key questions OR only <u>two</u> key questions referenced.	References <u>four</u> key questions.	References <u>six</u> key questions.	References all <u>eight</u> key questions.
No rationale provided for the applicability or inapplicability of any Key Questions to the ethical situation.	Provides a rationale for the applicability or inapplicability of <u>two</u> key questions to the ethical situation.	Provides a rationale for the applicability or inapplicability of <u>four</u> key questions to the ethical situation.	Provides a rationale for the applicability or inapplicability of <u>six</u> key questions to the ethical situation.	For all <u>eight</u> questions provides a rationale for its applicability or inapplicability to the ethical situation.
No attempt to analyze any <u>of the referenced</u> key questions.	Analysis attempted using two or more key questions. Typically <u>incorrect</u> ascription of the key questions to the ethical situation. Account is <u>unclear, disorganized, or inaccurate.</u>	Analysis attempted using three or more key questions. <u>Basically accurate</u> ascription of the key questions to the ethical situation. Account is <u>unclear</u> or <u>disorganized.</u>	Analysis attempted using three or more key questions. <u>Accurate</u> ascription of the key questions to the ethical situation. Account is <u>clear</u> and <u>organized.</u>	Meets criteria for *Excellent* AND… Nuanced treatment of key questions, for example: • elucidates subtle distinctions • uses analogies or metaphors • considers different issues within same key question
No judgment is presented OR judgment presented with no rationale.	Uses products of the analysis and provides some weighing to make a decision. Account is <u>unclear, disorganized, or inaccurate.</u>	Conveys weighing approach using analysis products. Provides an <u>intelligible</u> basis for judgment.	Meets criteria for *Good* AND…. Logically terminates in decision that will be reached.	Meets criteria for *Excellent* AND… Products of analysis weighed to make judgment <u>compelling.</u>

James Madison University © 2014

Appendix C

Ethical Reasoning Intervention Implementation Fidelity Checklist

Fidelity Researcher: _____   Date of Data Collection: _____

Targeted Objective for Learning Improvement	Intervention component	Duration in minutes (Actual)	Responsiveness 1 = Low (unengaged) 3 = Medium 5 = High (engaged)	Specific Features	Adherence Y/N	Quality 1 = Low (confusing) 3 = Medium 5 = High (clear)	Comments/ Additional Observations
Students will be able to apply their ethical reasoning skills to their personal, professional, and civic lives	Introduction/ Building Foundation to 8 Key Question (i.e., James Madison University's ethical reasoning framework)			Elaborate or unpack each of the 8 Key Questions ethical reasoning framework (e.g., reviewing the handbook, lecturing, PowerPoint slides, video clip, discussion)			
				Read/Review JMU's ethical reasoning student learning outcome			
				Read/Review rubric			
				Students experience a "check point" to check their own knowledge of the 8 Key Questions ethical reasoning framework (maybe use ethical reasoning content expert's multiple choice items??; crossword puzzle or word find; ball activity, news stories)			
				Map 8 Key Questions to some other work (can be something disciplinary like standards or something societal like policies or media or something practical, or something personal, news stories, onto class community or rules of engagement, etc.)			
				Critique/edit/comment/annotate the 8 Key Questions framework (e.g., could be wiki, could be collectively done in class, what do you like about 8 Key Questions? What would you change about them?; collective knowledge building)			
				Provide/discuss/present example of a decision-making process with AND without ethical reasoning ("ethical reasoning" is defined as being able to use 2+ Key Questions)			
Students will be able to apply their ethical reasoning skills to their personal, professional, and civic lives	Ethical Case Study			Review/Refresh 8 Key Questions ethical reasoning framework			
				Identify where/how each of the 8 Key Questions are/ are not applied within the case			
				Give/discuss rationale for how each of the 8 Key Questions are/are not applied			
				Engage in reflection (e.g., could be formal or informal, written, oral, group, what issues did you have, what was easy/hard)			
				Identify/discuss which (if any) aspects of the case are "compelling?" To what extent or degree was the case "compelling?"			

Appendix C, continued

Targeted Objective for Learning Improvement	Intervention component	Duration in minutes (Actual)	Responsiveness 1 = Low (unengaged) 3 = Medium 5 = High (engaged)	Specific Features	Adherence Y/N	Quality 1 = Low (confusing) 3 = Medium 5 = High (clear)	Comments/ Additional Observations
Students will be able to apply their ethical reasoning skills to their personal,	Examples			Have students together review/build a "strong" or "effective" example of ethical reasoning (e.g., show senior ethical reasoning faculty members students' videos in class and talk about what they could have done differently)			
professional, and civic lives				Identify and explain how characteristics or features make the case (in)effective referencing JMU's ethical reasoning student learning outcome and/or rubric?			
Students will be able to apply their ethical reasoning skills to their personal, professional, and civic lives	Multi-modal Analysis Visualization			Students experience (either visually or through some other sensory modality like touch, feel, movement, etc.) analysis processes- this can be "shown" by professor or created by students (e.g., block exercise, using color or size, show pre-made PowerPoint slides, students personify Key Question using their bodies as visuals, concept map- decision trees, Pictionary type game, role playing, collages, etc.)			
Students will be able to apply their ethical reasoning skills to their personal, professional, and civic lives	Analysis of 8 Key Questions and/or Analysis with 8 Key Questions			Students experience some sort of analysis (or breaking a part) of at least one Key Question; should get at nuances if possible			
				Identify obstacles or pitfalls to analysis (e.g., only analyzing 1 Key Question, confirmation bias, privilege)			
				Consider contextual factors (e.g., could include or "get at" multiple perspectives)			
				Expose/demonstrate/suggest how multiple perspectives can compete/interact w/one another within the same Key Question			
Students will be able to apply their ethical reasoning skills to their personal, professional, and civic lives	Weighing & Deciding using the 8 Key Questions as rationale			Students process something (debate, case, discussion, etc.) using the 8 Key Questions Students must arrive at or grapple with a particular conclusion or decision point			
				Multiple stakeholders and/or multiple perspectives are identified or considered			