Abstract

Higher education institutions struggle to demonstrate learning improvement (Banta, Jones, & Black, 2009; Banta & Blaich, 2011; Jankowski, Timmer, Kinzie, & Kuh, 2018). We showcase how student learning outcomes assessment processes can benefit from strong program theory and implementation fidelity data. In our example, faculty articulated the etiology of the distal outcome of acting ethically, which allowed for specification and measurement of the intermediate student learning outcomes. Faculty specified research-informed curriculum and pedagogy to influence the intermediate outcomes and ultimately the distal outcome. By articulating the program theory, faculty were able to assess both the intermediate outcomes for gains and their associated curriculum for implementation fidelity. Faculty could then identify what aspects of programming required changes to evidence learning improvement. Thus, we argue that program theory and implementation fidelity should be prominent components of higher education outcomes assessment processes to address the dearth of empirically supported learning improvement.

LEARNING **IMPROVEMENT**

AUTHORS

Kristen L. Smith, Ph.D.
Macmillan Learning

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

Elevating Program Theory and Implementation Fidelity in Higher Education: Modeling the Process via an Ethical Reasoning Curriculum

Improving student learning in higher education is challenging. Few universities or colleges have used outcomes assessment data to demonstrate learning improvement (Banta, et al., 2009; Banta & Blaich, 2011; Jankowski, et al., 2018; Kushimoto, 2010). Hence, assessment practitioners are critically reflecting on their practices and developing strategies to address this shortcoming (Coates, 2016; Fulcher & Prendergast, 2019; Fulcher et al., 2017; Mathers, et al., 2018; Smith et al., 2018). Expanding traditional outcomes assessment practices could increase the likelihood of positively impacting student learning. More specifically, by articulating strong program theory (Pope, et al. 2019) and collecting implementation fidelity data (Smith et al., 2019), faculty and student affairs practitioners should be able to identify what aspects of programming (i.e., educational interventions) require changes to achieve learning improvement.

In this paper, we describe how to incorporate strong program theory and implementation fidelity into assessment practice via five steps. To illustrate these steps, we provide an example of an ethical reasoning program at our institution. The processes we describe can be applied to any academic (e.g., Meixner et al., 2020) or student affairs (e.g., Fisher et al., 2014; Gerstner & Finney, 2013) educational program.

CORRESPONDENCE

Email
kristen.smith@macmillan.com

Expanding Higher Education Assessment Practice to Include Strong Program Theory

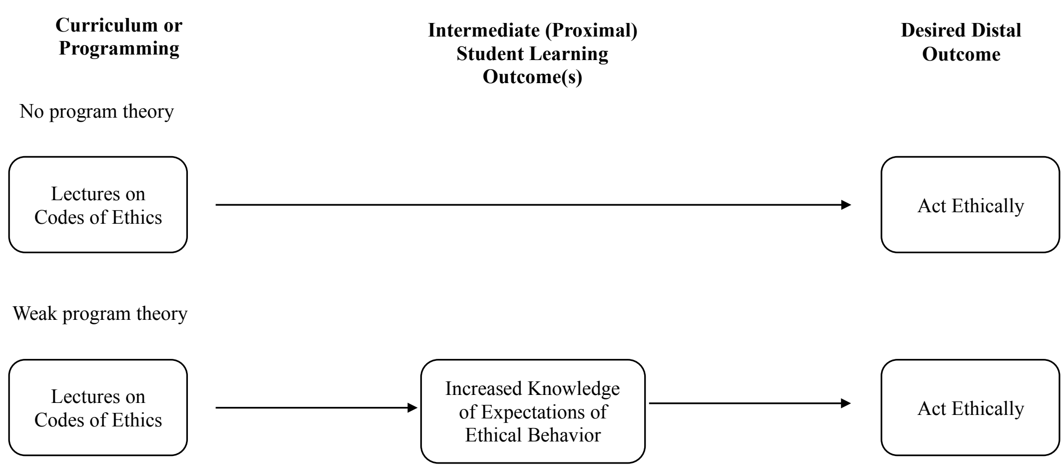
As faculty or student affairs practitioners, we are expected to design effective educational programs and assess their effectiveness (Coates, 2016; Finney & Horst, 2019a, 2019b; Leathwood & Phillips, 2000; U.S. Department of Education, 2006). However, the emphasis on gathering and reporting assessment data can distract from the equally important responsibility of designing intentional programming (e.g., curricula, activities) informed by theory and research. Creating such evidence-informed programming can be a daunting task, especially given the lack of practical guidance (Pope, et al., 2019). Yet, the use of theory and research to articulate strong program theory can inspire high-quality, valuable assessment practice.

The use of theory and research to articulate strong program theory can inspire high-quality, valuable assessment practice.

Program theory is defined as "the construction of a plausible and sensible model of how a program is supposed to work" (Bickman, 1987, p. 5). Program theory "clarifies the set of cause-and-effect relationships" believed to connect the things students do (i.e., programming) to the outcomes they are expected to achieve (Bickman, 1987, p. 5). Strong program theory is evidence-based and articulates coherent links between curriculum and/or pedagogies and student learning outcomes (SLOs). Weak program theory is often based on hunches, assumptions, or limited personal experiences.

For example, imagine if the faculty developing an ethical reasoning educational program were asked, "Why should this program result in the intended outcome of students acting ethically?" They may state, "We believe lectures on codes of conduct will increase students' knowledge regarding expectations of ethical behavior, and their increased knowledge will increase their ethical behavior." This statement would reflect their program theory (see Figure 1). However, without empirical evidence or established theory to support the link between knowledge of expectations of ethical behavior (intermediate outcome) and ethical behavior (distal outcome), the program theory would be weak.

Figure 1
Logic model depicting the difference between a program with no program theory and a program with weak program theory



In practice, we often observe weak program theory, which limits the use of assessment results to improve ineffective programs (Pope et al., 2019). In fact, we have witnessed rapid development of courses and programs based on hunches or beliefs even though established theory and empirical evidence could have informed course or program development. The most dire situation occurs when existing theory or research provides evidence *against* the hunches or beliefs used to guide program development.

To better integrate articulation of program theory and collection of implementation fidelity data into the assessment of educational programs, we guide readers through a five-step process (see Table 1). The process involves the following steps:

1. Articulate a feasible and malleable distal outcome;
2. Articulate theory- or research-based intermediate (proximal) outcomes;
3. Create intentional, theory- or research-based programming;
4. Collect implementation fidelity data to identify if the research-based programming was implemented;
5. Collect outcomes data to evaluate the effectiveness of the implemented programming.

Answering a series of questions associated with steps 1 to 3 facilitates building the program's logic (see Table 1). The resulting logic model clearly conveys "why" or "how" the programming should impact the distal outcome (see Figure 2). Given strong program theory

Table 1
Five-step process for building and evaluating a theory-based ethical reasoning program

Steps to Articulate & Evaluate Program Theory	Most Important Question(s) to Ask	Ethical Reasoning Example
1. Articulate the Distal Outcome	What problem or outcome needs attention? Is the outcome malleable?	The distal outcome is to significantly increase the frequency of ethical behaviors among students (e.g., students "act ethically").
2. Articulate Theory- or Research-Based Intermediate (Proximal) Student Learning Outcomes (SLOs)	What is the etiology (i.e., what are the causes) of the distal outcome based on current theory and research?	• A deliberative ethical reasoning process is needed to behave ethically. By engaging in a deliberative thought process, students avoid a quick, default, confirmatory decision regarding how to behave (Kahneman, 2011). • To develop this deliberative ethical reasoning process, students must be exposed to multiple considerations associated with an ethical decision or behavior. Traditional students are dualistic thinkers who consider decisions or behaviors as right or wrong (Perry, 1970); thus, we must expose them to multiple considerations. • Upon knowing multiple considerations (e.g., empathy, responsibility), students may tend to favor one consideration over others in most situations (e.g., Gilligan, 1982). Student must be challenged to wrestle with all considerations, which invokes cognitive dissonance and spurs growth in ethical reasoning skills (Gilligan, 1982: Kohlberg, 1981; Perry, 1970). • Given this research, the following intermediate SLOs were specified to influence the distal outcome of behaving ethically: o Students will state, from memory, the 8 Key Questions o Students will explain each Key Question o When given a specific decision and rationale on an ethical issue, students will identify the Key Question most consistent with the decision and rationale o For a hypothetical ethical dilemma, students will evaluate courses of action by applying (weighing & balancing) the Key Questions o In their own personal lives, students will evaluate courses of action based on a number of considerations (i.e., 8KQs)
3. Develop Theory-Based Programming/Curriculum to Impact the Intermediate SLOs	What programming affects the intermediate SLOs based on current theory and research?	• Program consisted of content and activities that foster the encoding, integration, and retention of information about the 8KQs. For example, students experienced at least one knowledge "check point" related to their understanding of the 8KQs, reviewed and refreshed the 8KQs within various case study/dilemma discussions, and mapped or represented the 8KQs in relation to some other work (e.g., disciplinary standards). These types of activities were supported by empirical research from the domain of cognition and learning (Halpern & Hakel, 2003). • Program also consisted of content and activities that utilized case study/dilemma discussions techniques and discipline-specific analysis of ethical cases, decisions, or dilemmas, given research suggested such techniques can promote ethical reasoning development (Bebeau; 1993; Keefer & Ashley, 2001; Wilhelm, 2010). • Program also consisted of numerous opportunities for guided reflection, as research suggested reflection is an important aspect of teaching ethics (Schmidt et al., 2009).
4. Collect Implementation Fidelity Data to Determine if Program Was Delivered as Intended	To what extent did the implemented or delivered program differ from the intended or planned program? Did students actually experience the programming?	• After articulating their program theory, faculty were able to create a fidelity checklist. • Creating the fidelity checklist facilitated the articulation and organization of specific curriculum features. • Reviewing the checklist before each class reminded faculty of the agreed upon and integral program features, guarding against program drift. • The fidelity checklist provided a systematic way to collect fidelity data and thus understand what version of the program students actually experienced.
5. Evaluate Outcomes Data to Inform Inferences about Program Effectiveness and Guide Changes in Program for Learning Improvement	Do assessment results suggest that the programming impacts the intermediate SLOs?	• Outcomes data were collected for the intermediate SLOs to assess change in students' ethical reasoning knowledge (e.g., constructed response, multiple-choice, and performance assessments administered before and after corresponding programming/curriculum). • Outcomes data were not collected for the distal outcome of acting ethically given this behavioral outcome is difficult to collect in real time. However, specification and assessment of theory- and research-based intermediate SLO's led faculty to believe that students achieving these intermediate outcomes are more likely to act ethically. • Fidelity and outcomes data were integrated to make more informed decisions about the program and better understand improvements in students' learning than afforded by outcomes data only.

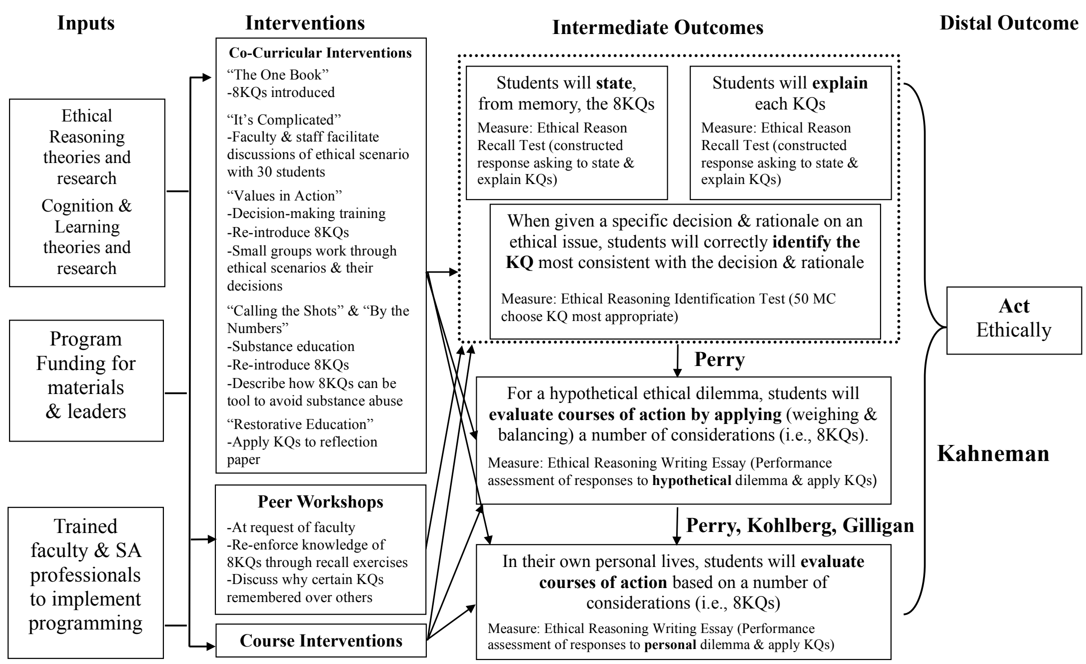
Note: The program theory incorporates theories and research related to ethical reasoning and moral development to specify and link the proximal intermediate outcomes to the distal outcome of acting ethically (i.e., Step 2). Using theories of learning and cognitive processing, the program theory also explicates how program components should affect the proximal intermediate outcomes (i.e., Step 3).

is articulated, practitioners can then empirically evaluate the theory-based programming (see Table 1, Steps 4 and 5).

Using an ethical reasoning program from our campus, we model a five-step process to articulate strong program theory and assess program effectiveness with regards to learning improvement. Although our example is complex, involving multiple intermediate outcomes and faculty across the institution, the five-step process can be applied to a variety of learning outcomes and educational programming (e.g., Fisher et al., 2014; Gerstner & Finney, 2013; Meixner et al., 2020; Pope et al., 2019).

Figure 2

Example program theory for an ethical reasoning intervention



Step 1: Articulate a malleable distal outcome. Creating a theory- or research-based educational program begins by specifying the ultimate or "distal" outcome one hopes to achieve. For the current example, this distal outcome was a result of our institution's Quality Enhancement Plan (QEP) for accreditation through the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC).

Before selecting ethical behavior as the distal outcome, university stakeholders asked: "Is it theoretically possible to impact ethical behavior in a college student population?" "Is acting ethically a *malleable* behavior?" "Can ethical behavior be learned?" If ethical behavior is stable or trait-like, developing programming to try to increase it would be a waste of university resources.

"Is acting ethically a malleable behavior?" "Can ethical behavior be learned?" If ethical behavior is stable or trait-like, developing programming to try to increase it would be a waste of university resources.

Informed by research, university stakeholders deemed ethical behavior as malleable and they understood intentional instruction would be necessary to build reasoning strategies to influence ethical behavior (Sanchez et al., 2017). For example, Keller (2010) defined ethics as something that can be practiced through "applied methods of rational inquiry to moral problems" (p. 12), suggesting ethical behavior can be impacted by particular approaches. Similarly, research in cognitive psychology provided evidence that many everyday behaviors result from fast, intuitive, or "gut" responses (Kahneman, 2011), which can be interrupted and slowed by the introduction of a prompt or thinking strategy (Ariely, 2013). Reasoning strategies can influence ethical behavior, but these strategies do not develop due to maturation alone. Instead, progression from basic to more advanced stages of reasoning requires effortful development (Kohlberg, 1969; Kohlberg, 1977).

Next, university stakeholders asked, "Given ethical behavior is malleable, can we *impact* it within the time and resource constraints of a traditional, four-year college experience?" Previous research suggested they could. Since the 1970s, several studies have linked participation in college to ethical reasoning development (King & Mayhew, 2002; Pascarella & Terenzini, 1991; Pascarella & Terenzini, 2005; Rest, 1979; Rest et al., 1986; Rest & Thoma, 1985). Co-curricular and classroom-based experiences have had a significant, but small effect on college students' ethical reasoning, especially for first-year students (Mayhew & Engberg, 2010; Mayhew, et al., 2010).

On our campus, when university stakeholders addressed Step 1, research and theory determined that ethical reasoning was a malleable and feasible outcome to target. Although research suggested that students' ethical reasoning behaviors could be impacted within the context of a college experience (King & Mayhew, 2002; Pascarella & Terenzini, 1991; Pascarella & Terenzini, 2005; Rest, 1979; Rest et al., 1986; Rest & Thoma, 1985), ethical behavior is a complex outcome. University stakeholders anticipated that this outcome may not be realized due to a single program, course, or intervention. Different types of interventions would likely impact different causes of ethical behavior. Thus, during Step 2 of the process, it would be critical for faculty and stakeholders to specify the more proximal, intermediate student learning outcomes that would influence the ultimate, distal outcome of ethical behavior (i.e., "acting ethically").

Research suggested that students need a deliberative ethical reasoning process to behave ethically.

Step 2: Specify theory- or research-based intermediate outcomes. Once the distal outcome of ethical behavior was specified in Step 1, the next step was to consult relevant empirical research and theory to articulate the underlying causes or influences of the behavior. Step 2 is difficult, time consuming, and critically important as it specifies what student characteristics must be influenced to achieve the distal outcome. We provide a detailed description of the process followed on our campus in order to support others engaging in this step.

Faculty tasked with creating the program's curriculum must understand the etiology of acting ethically (West & Aiken, 1997). From this understanding, *intermediate* SLOs were specified (see Table 1 and Figure 2). These intermediate SLOs answered the question: "What do students need to *know*, *feel* or *perceive* (i.e., attitudes), and *do* (i.e., skills) to achieve the distal outcome of acting ethically?" Research suggested that students need a deliberative ethical reasoning process to behave ethically. By engaging in a deliberate ethical reasoning thought process, students can avoid a quick, default, confirmatory decision regarding how to behave (Kahneman, 2011).

Helping students engage in a deliberative thought process can be challenging given students' thought processes are naturally automated, rapid, and rooted in intuitive or "gut" reactions. Thus, students need a strategy or process to help slow their default thinking and instead engage in a deliberative reasoning process (Ariely, 2013). To provide students with such a strategy, stakeholders created the "8 Key Questions" (i.e., the 8KQs) ethical reasoning framework (Sanchez et al., 2017). This deliberative ethical reasoning process prompts students to consider, weigh, and balance the following Key Questions when grappling with an ethical issue or dilemma:

1. Fairness: *How can I (we) act justly, equitably, and balance legitimate interests?*
2. Outcomes: *What possible actions achieve the best short- and long-term outcomes for me and all others?*
3. Responsibilities: *What duties and/or obligations apply?*
4. Character: *What actions help me (us) become my (our) ideal self (selves)?*
5. Liberty: *How do I (we) show respect for personal freedom, autonomy, and consent?*
6. Empathy: *How would I (we) act if I (we) cared about all involved?*
7. Authority: *What do legitimate authorities (e.g., experts, law, my religion/god) expect?*
8. Rights: *What rights, if any, (e.g., innate, legal, social) apply?*

The 8KQs incorporated ideas from the following philosophical perspectives: John Stuart Mill's Utilitarian theory, Kant's natural duties and obligations, Rawls' justice as fairness, Kohlberg's role of authority, Gilligan's role of empathy, and Aristotle's virtuous self (Lehnen & Pyle, 2019).

Committing the 8KQs to memory and being able to explain them were considered necessary (but not sufficient) to acting ethically. That is, the deliberative ethical reasoning process is unpacked in Figure 2 as five intermediate student learning outcomes, with memorizing and explaining the 8KQs being foundational knowledge necessary to engage in the process of ethical reasoning.

Beyond being able to *state* and *explain* the 8KQs, students need to be able to recognize which considerations are being applied by others. As shown in Figure 2, another intermediate outcome involves students being able to *identify* which Key Question is most consistent with a given ethical decision and rationale. Traditional students are dualistic thinkers who tend to consider decisions or behaviors as right or wrong (Perry, 1970). The 8KQs framework intentionally exposes students to multiple considerations associated with an ethical dilemma. Students should understand that a particular ethical dilemma can be associated with any consideration (e.g., Fairness, Authority), and this consideration likely influences one's subsequent behavior.

Upon knowing multiple considerations (i.e., 8KQs), students may tend to favor one consideration over others (Gilligan, 1982). Moreover, stakeholders acknowledged that simply being able to recall, explain, and identify considerations (i.e., KQs) associated with a dilemma (i.e., the first three intermediate SLOs in logic model in Figure 2) would not be sufficient for students to achieve the distal outcome of acting ethically. Thus, two additional, application-focused intermediate SLOs were articulated: students should *evaluate* courses of action by *applying* the 8KQs to hypothetical ethical dilemmas and to personal ethical dilemmas in their own lives (Lehnen & Pyle, 2019). Application of the 8KQ involves weighing the applicability of the considerations raised by each KQ, given the context of the ethical dilemma, and appropriately balancing those considerations to make a conclusion or grapple with a decision. To apply the 8KQs, students must wrestle with all considerations (e.g., 8KQs), which should invoke cognitive dissonance or disequilibrium and spur growth in ethical reasoning skills (Gilligan, 1982; Kohlberg, 1981; Perry, 1970; Schmidt et al., 2009).

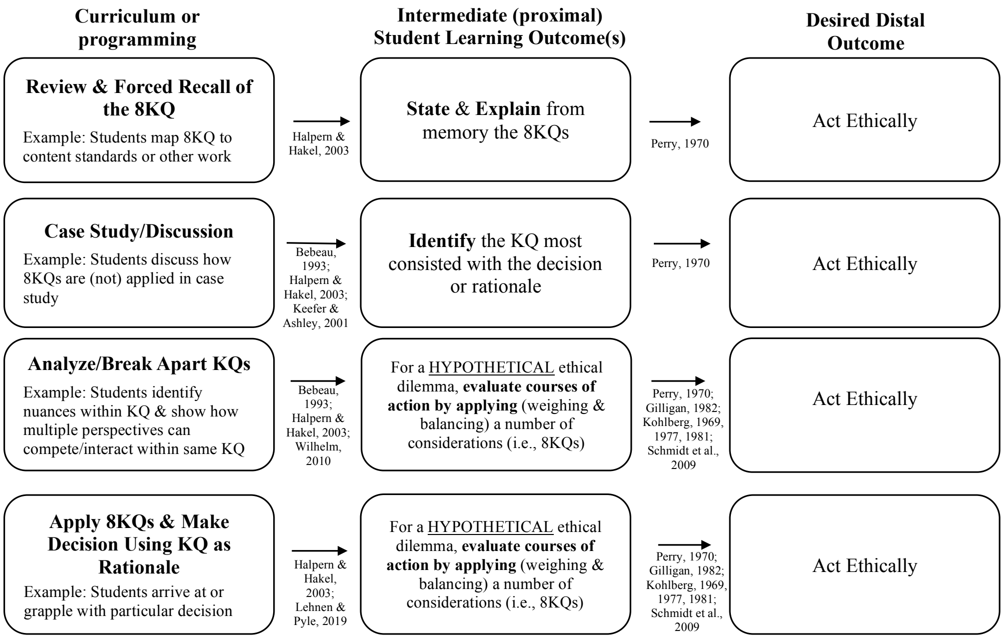
The etiology of acting ethically was articulated in Step 2. The specific etiology articulated in Step 2 will vary depending on the distal outcome of interest. Regardless of the outcome, Step 2 involves using theory and research to convey the underlying causes of the desired distal outcome and articulating these in terms of intermediate student learning outcomes.

Regardless of the outcome, Step 2 involves using theory and research to convey the underlying causes of the desired distal outcome and articulating these in terms of intermediate student learning outcomes.

For example, using theory and research, stakeholders (including experts in ethical reasoning) indicated that being able to state, explain, and identify the 8KQs facilitates students being able to apply the 8KQs within the contexts of hypothetical and/or personal ethical dilemmas. This application forces students to practice complex thought processes (Lehnen & Pyle, 2019). Being challenged to consider alternative perspectives (i.e., 8KQs) and appropriately weigh and balance these different perspectives (i.e., apply the 8KQs) enables students to act ethically (the desired distal outcome) when they are confronted with ethical dilemmas (Sanchez et al., 2017). These various intermediate SLOs would likely be impacted by different kinds of programming. Thus, during Step 3 of the process, it would be crucial for faculty and stakeholders to specify research-informed programming (i.e., curricula and pedagogies) that would positively influence the intermediate SLOs.

Step 3: Develop theory- or research-based programming aligned to intermediate outcomes. Once the distal outcome of ethical behavior and intermediate SLOs that influence ethical behavior were specified (i.e., Step 1 and 2, respectively), the next step was to determine how to achieve the five intermediate SLOs via programming (e.g., activities, curriculum, pedagogies). At this step, stakeholders asked, "Given achievement of these intermediate outcomes should increase the likelihood of our students acting ethically (i.e., the distal outcome), how can faculty and practitioners *intervene* to support students achieving these intermediate outcomes?" "What *curricular* or *pedagogical strategies* do research or theory suggest may be effective to influence the intermediate outcomes?"

Figure 3
Logic model depicting example curriculum features that were evidence-based resulting in strong program theory



Note: “Hypothetical” ethical dilemmas can be interchanged with “Personal” ethical dilemmas in this logic model.

Just as research and theory informed the articulation of the five intermediate SLOs and the link between them and the distal outcome of ethical behavior, research and theory informed the programming (i.e., curricular, pedagogical components) and linked programming to the intermediate SLOs (see Figure 3). Recall, program theory involves undergirding each arrow in a logic model with research or theory that supports the links (Baldwin et al., 2004). Figure 2 includes several of the interventions built to influence the intermediate SLOs. Some programming was experienced by all students as part of mandatory summer orientation for first-year, entering students. Other programming was experienced by a smaller number of students (e.g., substance abuse education).

Articulating a strong program theory and creating research-informed programming is not sufficient to achieve the SLOs. Students must actually experience the programming in order to achieve the intermediate and distal outcomes.

For the purpose of this paper, we describe the longest intervention— a semester-long ethical reasoning curriculum created by faculty and infused within six cross-disciplinary courses (see Table 2). For the ethical reasoning curriculum, the faculty first examined research that evaluated the effectiveness of particular activities to influence the intermediate SLOs (see Table 1). They consulted research examining best strategies for learning, retention, and future application (e.g., Fink, 2013). They also shared and discussed previous approaches to teaching relevant concepts and identified if those approaches were evidence-based (Smith et al., 2017).

Using this process, faculty co-created a research-informed curriculum that could be implemented across the variety of content domains (see Figure 3). For example, research suggests that practice at retrieval spaced over time promotes long-term retention (Halpern & Hakel, 2003); thus, the faculty designed the ethical reasoning curriculum to include “Review and Forced Recall” activities that would support the encoding and retention of the 8KQs into long-term memory. Students experienced knowledge “check points” of their ability to state and explain the 8KQs. Moreover, encoding of information can be improved when students are asked to “re-represent” information in an alternative format (Halpern & Hakel, 2003). Thus, faculty included a program feature that asked students to map the 8KQs to disciplinary content or other areas of interest (e.g., industry standards, policies of practice, news stories, media).

Research suggests that “Case Study/Discussion” techniques promote ethical reasoning development (Bebeau, 1993). Case-based approaches to teaching ethics have pedagogical utility because they provide opportunities for students to discuss and disagree (Keefer & Ashley, 2001). Thus, the faculty asked students to identify and discuss the (in)applicability of each of the 8KQs within a given case study or ethical dilemma (see Figure 3). They also asked students to identify aspects of case studies that were compelling in relation to the 8KQs. Furthermore, opportunities for guided reflection are important aspects of teaching ethics (Schmidt et al., 2009). Thus, faculty had students engage in reflection about ethical case studies (e.g., what aspects of case studies were compelling) in formal and informal ways (e.g., oral, written, group, individual).

Table 2
Description of six cross-disciplinary, semester-long courses in which faculty infused theory-based ethical reasoning programming

# of Students	Domain/Discipline	Brief Description of Course	Course Type
77	Health Sciences	Upper level students; Required course for major; Ethics in class title	Lecture
18	Philosophy	Lower level students; General Education Class; Fulfills Cluster 1 requirement; Ethics in class title	Lecture
7	Justice Studies	Upper level students; Elective Course	Seminar; Community Service Learning
7	Integrated Science & Technology	Upper level students; Elective Course	Seminar; Community Service Learning
42	Education	Upper level students; Course for minor	Lecture; Community Service Learning
40	Health Sciences	Upper level students; Required course for major	Lecture

Coupling the implementation fidelity and outcomes assessment data allowed the faculty to understand variability in students' ethical reasoning skills given differences in the extent to which the ethical reasoning programming was implemented with high fidelity.

Research indicates the importance of discipline-specific analysis and examination of ethical issues that students may actually face (Bebeau, 1993; Wilhelm, 2010). Therefore, faculty asked students to "Analyze/Break Apart KQs" relevant to ethical dilemmas within the contexts of their own disciplines (see Figure 3). Students were also asked to grapple with multiple perspectives – within the same KQ – that may compete, interact, or disagree.

Lehnen and Pyle (2019) suggested that students must be challenged to move forward in their ethical reasoning processes and behaviors through interacting with fictional and real-life ethical dilemmas. As students make decisions regarding ethical dilemmas, they will likely experience cognitive dissonance which can spur growth in ethical reasoning skills (Gilligan, 1982; Kohlberg, 1981; Perry, 1970; Schmidt et al., 2009). Therefore, faculty asked students to "Apply the 8KQs and Make Decisions Using KQ as their Rationale" (see Figure 3). For example, students were asked to grapple with a particular ethical decision that someone else made and/or arrive at their own ethical decision using the 8KQ. Faculty also asked students to consider multiple stakeholders and/or perspectives when applying the considerations raised by the 8KQ.

Clearly, the faculty invested substantial time and effort in Step 3 to determine what programming should influence the intermediate SLOs. As outlined in Step 3, development of programming was informed by research, theory, and previous teaching experiences. However, as any instructor knows, articulating a strong program theory and creating research-informed programming is not sufficient to achieve the SLOs. Students must actually *experience* the programming in order to achieve the intermediate and distal outcomes. For students to maximally benefit from research-informed programming, classroom implementation has to be considered (Little & Hahs-Vaugh, 2007). In short, high quality implementation is a necessary aspect of effective programming (Durlak, 2016). Thus, the faculty wanted to empirically evaluate the extent to which the new programming was actually implemented.

Expanding Higher Education Assessment Practice to Include Implementation Fidelity

Implementation fidelity data allow faculty to determine the extent to which the programming *as designed* differs from the programming *as delivered* (Gerstner & Finney, 2013; O'Donnell, 2008; Smith et al., 2017; Smith et al., 2019). Fidelity data allow stakeholders to better understand the (in)effectiveness of specific features of the educational intervention (Cook & Shadish, 1986) and, in turn, make appropriate modifications (Finney & Smith, 2016). However, traditional outcomes assessment approaches (e.g., Walvoord, 2010) exclude collection of data reflecting the alignment between the *designed* and *delivered* programming. Thus, assessment practice should be expanded to include the collection and use of implementation fidelity data after program theory has been articulated.

Step 4: Collect implementation fidelity data to determine if program was delivered as intended. After faculty co-created the research-based programming aligned with the intermediate outcomes (i.e., Step 3), they needed to determine if the programming was delivered as intended. Thus, they created an implementation fidelity checklist and used it to capture data concerning four aspects of implementation fidelity: 1) whether each program feature was delivered; 2) the quality with which each feature was delivered; 3) the perceived student responsiveness or engagement during a given feature; and 4) the duration of time for each feature (Gerstner & Finney, 2013). See Smith et al. 2017 for more information about the fidelity checklist used by the faculty on our campus.

Implementation fidelity data provided faculty an opportunity to explore the relative effectiveness of specific features of the ethical reasoning programming that they invested a great deal of time and effort co-creating.

Either one or two trained implementation fidelity researchers attended live class sessions throughout the semester and used the checklist to collect fidelity data. Faculty members were asked to review data collected by the fidelity researchers (e.g., to note anything that may have been missed or misrepresented). In addition, for at least three class sessions, each faculty member filled out the checklist as a "self-audit" indication of fidelity (Smith et al. 2017).

In general, the ethical reasoning curriculum was implemented with high fidelity (i.e., strong alignment between planned and experienced programming) because the faculty understood that the influence of their research-informed programming on the SLOs was moderated by implementation fidelity. Articulating a strong program theory enabled the faculty to create a useful implementation fidelity checklist. Creating the fidelity checklist also helped faculty articulate the specific curriculum features. Reviewing the checklist before each class reminded faculty of the agreed upon program features, guarding against program drift. At the end of the semester, faculty commented that using the fidelity checklist added structure to their teaching. The fidelity checklist allowed them to plan their ethical reasoning course materials with greater precision (Smith et al., 2017).

Given the time and resources spent developing the research-informed ethical reasoning programming (i.e., Step 3), the faculty were genuinely excited to assess the extent to which that programming was implemented (i.e., Step 4) and determine if programming was associated with student achievement of the intermediate SLOs (i.e., Step 5). Coupling the implementation fidelity and outcomes assessment data allowed the faculty to understand variability in students' ethical reasoning skills given differences in the extent to which the ethical reasoning programming was implemented with high fidelity. That is, implementation fidelity data provided faculty an opportunity to explore the relative effectiveness of specific features of the ethical reasoning programming that they invested a great deal of time and effort co-creating.

A Call for "Expanded" Assessment Practice in Higher Education

Our experience expanding assessment practice to include strong program theory and implementation fidelity yielded positive results. The combination of program theory, implementation fidelity, and well-aligned outcomes assessment instruments provided:

1. an understanding of *why* students' skills improved over time (e.g., which aspects of the research-informed programming appeared to positively influence students' learning);

2. information to make informed modifications to the programming; and

3. evidence of effective program features that could be shared with colleagues interested in improving similar intermediate or distal learning outcomes.

Had faculty followed a more traditional assessment cycle (e.g., Walvoord, 2010), they *may not* have articulated a research-based program, and they *would not* have collected implementation fidelity data. Had the program theory not been articulated and only the distal outcome of ethical behavior been assessed, how could these limited data be used for program improvement? Had implementation fidelity data not been collected, how could faculty link aspects of programming to improvements in students' learning?

We believe outcomes assessment data have limited utility and thus should not be collected until stakeholders can answer two basic questions: "Why should this programming result in the intended outcome?" (i.e., program theory) and "Was the research-informed programming actually experienced by students?" (i.e., implementation fidelity).

Without strong program theory and implementation fidelity, it is difficult, if not impossible, to determine which intermediate outcomes are achieved and which program features are effective. By specifying how the different program features should (based on research) result in achievement of the intermediate SLOs and how the intermediate SLOs should (based on research) help students progress toward the distal SLO, faculty are able to collect the data necessary to make valid inferences about program effectiveness. Moreover, they can share those results and the new programming with colleagues.

Despite the positive effects of expanding the traditional assessment cycle, there are challenges. For example, a program theory may have been developed to explicate the logic of activities, but the theory was never intentionally communicated (Leeuw, 2003). Thus, new faculty or facilitators may engage in an unnecessary program overhaul because they are not privy to the program's logic. Time must be allocated to create a record of the development of program theory in order to reap the benefits of this difficult, yet critical work. A greater challenge is that faculty or program facilitators may struggle to articulate connections between outcomes and actions (Savaya & Waysman, 2005). Thus, educational interventions are developed (and assessed) without a clear theory or evidence base as a foundation (Bickman, 1987). This challenge stems from the paucity of methods that describe how to do so and lack of training in this domain (Leeuw, 2003; Pope et al., 2019). Similarly, there is a lack of didactic guidance regarding implementation fidelity processes (O'Donnell, 2008; Smith et al., 2017). Collecting implementation fidelity data can be logistically challenging and resource intensive. Moreover, faculty or program facilitators must be willing to have their programming observed, recorded, or otherwise "audited."

Acknowledging these challenges, we urge faculty and practitioners to expand their assessment processes to include the explicit articulation of strong program theory and collection of implementation fidelity data. We have didactically outlined five steps to build and evaluate an evidence-based program that should be effective (see Table 1). We believe outcomes assessment data have limited utility and thus should not be collected until stakeholders can answer two basic questions: "Why should this programming result in the intended outcome?" (i.e., program theory) and "Was the research-informed programming actually experienced by students?" (i.e., implementation fidelity). Although some may find that assertion extreme, it is only after program theory has been articulated that faculty can collect relevant outcomes data. Moreover, valid inferences from outcomes data are contingent on understanding what programming the students actually experienced. This "expanded" assessment practice has great potential to provide better-designed, more effective, research-informed programming. As students have opportunities to experience well-implemented, research-informed programming, their learning should demonstrably improve.

References

Ariely, D. (2013). *The honest truth about dishonesty: How we lie to everyone—especially ourselves*. New York, NY: HarperCollins.

Baldwin, C. K., Hutchinson, S. L., & Magnuson, D. R. (2004). Program theory: A framework for theory-driven programming and evaluation. *Therapeutic Recreation Journal, 38*, 16-31.

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

Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice*. San Francisco, CA: Jossey-Bass.

Bebeau, M. J. (1993). Designing an outcome-based ethics curriculum for professional education: Strategies and evidence. *Journal of Moral Education, 22*, 313-326.

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation, 33*, 5-18.

Coates, H. (2016). Assessing student learning outcomes internationally: Insights and frontiers. *Assessment & Evaluation in Higher Education, 41*, 662–676.

Cook, T. D., & Shadish, W. R. (1986). Program evaluation: The worldly science. *Annual Review of Psychology, 37*, 193-232.

Durlak, J. A. (2016) Programme implementation in social and emotional learning: Basic issues and research findings. *Cambridge Journal of Education, 46*, 333-345.

Fink. L. D. (2003). *Creating significant learning experiences: An integrated approach to designing college courses*. San Francisco, CA: Jossey-Bass.

Finney, S. J., & Horst, S. J. (2019a). Standards, standards, standards: Mapping professional standards for outcomes assessment to assessment practice. *Journal of Student Affairs Research and Practice, 56*, 310-325.

Finney, S. J., & Horst, S. J. (2019b). The status of assessment, evaluation, and research in student affairs. In V. L. Wise & Z. Davenport (Eds.), *Student affairs assessment, evaluation, and research: A guidebook for graduate students and new professionals* (pp. 3-19). Springfield, IL: Charles Thomas Publisher.

Finney, S. J., & Smith, K. L. (2016). *Ignorance is not bliss: Implementation fidelity and learning improvement*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

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., & Prendergast, C. O. (2019). Lots of assessment, little improvement? How to fix the broken system. In S. P. Hundley & S. Kahn (Eds.), *Trends in assessment: Ideas, opportunities, and issues for higher education*. Sterling, VA: Stylus.

Fulcher K. H., Smith, K. L., Sanchez, E. R. H., Ames, A. J., & Meixner, C. (2017). Return of the pig: Standards for learning improvement. *Research & Practice in Assessment, 11*, 10-27.

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

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning. *Change: The Magazine of Higher Learning, 35*, 36-41.

Jankowski, N. A., Timmer, J. D., Kinzie, J., & Kuh, G. D. (2018). *Assessment that matters: Trending toward practices that document authentic student learning*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Keefer, M., & Ashley, K. (2001). Case-based approaches to professional ethics: A systematic comparison of students' and ethicists' moral reasoning. *Journal of Moral Education, 30*(4), 377-398.

Keller, D. R. (2010). An introduction to ethics for teaching. *Teaching Ethics, 11*, 1-54.

King, P. M., & Mayhew, M. J. (2002). Moral judgment development in higher education: Insights from the defining issues test. *Journal of Moral Education, 31*, 247-270.

Kohlberg, L. (1969). Stage and sequence: The cognitive approach to socialization. In D.A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago, IL: Rand McNally.

Kohlberg, L. (1977). Implications of moral stages for adult education. *Religious Education, 72*, 183-201.

Kohlberg, L. (1981). *Essays on moral development: The philosophy of moral development* (Vol. 1). San Francisco, CA: Harper & Row.

Kushimoto, T. (2010). Outcomes assessment and its role in self-reviews of undergraduate education: In the context of Japanese higher education reforms since the 1990s. *Higher Education, 59*, 589-598.

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation, 24*(1), 5-20.

Lehnen, A. R., & Pyle, L. K. (2019). Ethical and moral reasoning development across campus. In K. Branch, J. S. Hart-Steffes, & C. M. Wilson (Eds.), *Applying student development theories holistically: Exemplar programming in higher education* (pp. 79-93). New York, NY: Routledge.

Leathwood, C., & Phillips, D. (2000). Developing curriculum evaluation research in higher education: Process, politics and practicalities. *Higher Education, 40*, 313-330.

Little, M. E, & Hahs-Vaughn, D. L. (2007). The implementation of content enhancement routines for improved content literacy for middle and secondary social studies students. *Journal of Personnel Evaluation in Education, 20*, 261-280.

Mathers, C.E., Finney, S. J., & Hathcoat, J. (2018). Student learning in higher education: A longitudinal analysis and faculty discussion. *Assessment and Evaluation in Higher Education, 43*, 1211-1227.

Mayhew, M. J., & Engberg, M. E. (2010). Diversity and moral reasoning: How negative diverse peer interactions affect the development of moral reasoning in undergraduate students. *The Journal of Higher Education, 81*(4), 459-488.

Mayhew, M. J., Seifert, T. A., & Pascarella, E. T. (2010). A multi-institutional assessment of moral reasoning development among first-year students. *The Review of Higher Education, 33*(3), 357–390.

Meixner, C., Pope, A. M., & Horst, S. J. (2020). Implementation fidelity in the classroom: Exploring the alignment of pedagogy to learning outcomes. *The Journal of Faculty Development, 34*, 15-22.

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*, 33-84.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco, CA: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students*. San Francisco, CA: Jossey-Bass.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York, NY: Holt, Rinehart, & Winston.

Pope, A., Finney, S. J., & Bare, A. (2019). The essential role of program theory: Fostering theory-driven practice and high-quality outcomes assessment in student affairs. *Research & Practice in Assessment, 14*, 5-17.

Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis, MN: University of Minnesota Press.

Rest, J. R., Deemer, D., Barnett, R., Spickelmier, J., & Volker, J. (1986). Life experiences and developmental pathways. In J. Rest (Ed.), *Moral development: Advances in research and theory* (p. 28-58). New York, NY: Praeger.

Rest, J. R., & Thoma, S. (1985). Relationship of moral judgment development to formal education, *Developmental Psychology, 21*, 709-714.

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*, 30-36.

Savaya, R., & Waysman, M. (2005). The logic model: A tool for incorporating theory in development and evaluation of programs. *Administration in Social Work, 29*(2), 85-103.

Schmidt, C., McAdams, C., & Foster, V. (2009). Promoting the moral reasoning of undergraduate business students through a deliberate psychological education-based classroom intervention. *Journal of Moral Education, 38*, 315-334.

Smith, K. L., Finney, S. J., & Fulcher, K. H. (2017). Actionable steps for engaging assessment practitioners and faculty in implementation fidelity research. *Research & Practice in Assessment, 12*, 71-86.

Smith, K. L., Finney, S. J., & Fulcher, K. H. (2019). Connecting assessment practices with curricula and pedagogy via implementation fidelity data. *Assessment and Evaluation in Higher Education, 44*, 263–282.

Smith, K. L., Good, M. R., & Jankowski, N. (2018). Considerations and resources for the learning improvement facilitator. *Research & Practice in Assessment, 13*, 20-26.

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education* (ED Pubs No. ED002591P).

Walvoord, B. E. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed.). San Francisco, CA: Jossey-Bass Publishers.

West, S., & Aiken, L. (1997). Toward understanding individual effects in multicomponent prevention programs: Design and analysis strategies. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 167–209). Washington, DC: American Psychological Association.

Wilhelm, W. (2010). Ethical reasoning instruction in non-ethics business courses: A nonintrusive approach. *The Delta Pi Epsilon Journal, 52*, 152-167.