What’s Your Theory of Action? Making Good Trouble With Literacy Assessment

David Slomp, Norbert Elliot

By their very nature, all assessments are consequential. Assessment programs and the measures that reside within them are designed to help teachers, administrators, and policymakers draw inferences about student ability. In turn, these test-based inferences inform decisions that impact individuals, classrooms, schools, and educational systems.

Although contemporary methods of validation are designed to examine consequences derived from the interpretation and use of test scores, they are not equipped to monitor how assessments alter—intentionally or otherwise—student, teacher, and system-level behaviors (Koretz & Hamilton, 2006; Slomp, 2020). In an era of test-based accountability—where assessments are explicitly intended to shape educational systems and to steer those working within them toward defined outcomes—the need to evaluate the broad set of consequences stemming from an assessment’s design and use is especially pronounced.

In this column, we present two related approaches—a Theory of Action (ToA) and an Integrated Design and Appraisal Framework (IDAF)—that can be leveraged by teachers to draw attention of test designers, policy leaders, and the public to the broad set of test components and consequences stemming from an assessment’s implementation. We demonstrate the force of an integrated ToA and IDAF approach through a case study of a required secondary school English Language Arts Diploma Examination, and we conclude with implications for classroom teachers. Our aim in this column is to advance a way for teachers to leverage power in literacy assessments on behalf of their students by raising questions regarding the components of an assessment, its delivery mechanisms, and its manifold consequences.

Theory of Action

Often considered a founder of social psychology, Kurt Lewin (1946) led the development of action research. A proponent of field work, Lewin developed an empirical investigative approach in the mid-1940s that focused on the very stakeholders who would most likely feel the consequences of any action involving them. His insights—emphasis on detailed fact finding, attitudes toward inquiry, encouragement of participatory communities, identification of power structures, and claims tempered by qualification—provide the impetus for theorizing assessment policy through the lens of action research.

By 1949, Stephen Corey of Columbia University had turned Lewin’s social policy lens to educational research. Noting that educational research was often conducted by those external to the very places—the classrooms and schools—where learning occurred, Corey (1949) emphasized the role of the teacher: “The action researcher works in a specific, dynamic situation with specific and identifiable persons” (p. 512), toward the aim of achieving “possible improvements in practice” (p. 513; see also Corey, 1953). Focusing on relationships between organizational structures and individuals, Chris Argyris (1957), a student of Lewin’s, turned his attention in 1957 to “to derive specific hypotheses regarding their mutual impact” (p. 2). Working with philosopher Donald Schön, in 1974 Argyris created what became known as a ToA framework: identification of program elements (the governing variables of a proposed learning intervention), hypotheses regarding action mechanisms (implementation strategies for the intervention), and potential consequences (the intermediate and long-term negative and positive impact that will likely arise from the intervention; Argyris &
Schön, 1974; see also Argyris, 1997). Today, ToA statements are recommended by the Institute of Education Sciences and the National Science Foundation (2013) to accompany early-stage, exploratory research and to inform planned interventions. Such statements, as these federal agencies have realized, transform the too often abstract nature of educational research into actionable agendas with real consequences.

Integrated Design and Appraisal Framework
Compared with ToA, IDAF has a much later origin, in 2016, and was developed with the specific purpose of promoting fairness in writing assessment. Informed by Messick’s (1980) work on the social consequences of testing, Kane’s (2013) evidence-based approach to score interpretation and use, and the renewed emphasis on fairness in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), IDAF was designed to enable literacy educators to pay systematic attention to the broad set of consequences derived from an assessment’s design and use. An IDAF analysis is enacted in the six phases shown in Figure 1 (for a more complete discussion of this process, see Slomp, 2016).

Case Study: Alberta’s Diploma Examination Program
To avoid working in generalities, we offer an example of our integrated ToA and IDAF approach. We have

Figure 1
Phases and Key Questions of the Integrated Design and Appraisal Framework

- **Phase 1: Identify Aims**
  - What purpose is the proposed assessment designed to achieve?
  - What are the potential consequences that could result from the implementation of this assessment program?
  - Who are the stakeholders who will be impacted by this proposed assessment?
  - How will the development and use of this assessment promote equity of opportunity for impacted populations?

- **Phase 2: Identify Target Domain**
  - What constructs are targeted in the assessment?
  - Does narrowing the construct to its sampling in the assessment either imperil achievement of intended outcomes or introduce potential unintended outcomes?

- **Phase 3: Analyze Assessment Design**
  - How completely do assessment items capture the construct sample?
  - In what ways do the assessment items introduce construct irrelevant variance to the assessment data?
  - How will the assessment framework and item design impact achievement of intended outcomes?
  - How will the assessment framework and item design support equity and fairness?
  - How can score reports be designed to support the intended outcomes?

- **Phase 4: Analyze Scoring System**
  - How do the scoring criteria ensure construct coverage?
  - How will scoring procedures support the achievement of intended outcomes?
  - How will scoring procedures support equity and fairness?
  - What unintended outcomes might result from implementation of the scoring procedures?

- **Phase 5: Analyze Assessment Results**
  - Has each link in the chain of inferences (scoring, generalization, extrapolation, decision) been tested for each intended use and has each link been supported by empirical evidence?
  - Within the population being assessed, have the populations who might be disparately impacted by the assessment been identified?
  - In cases where differences in performance across populations have been identified, does evidence based on test content, test context, test response, and opportunity to learn indicate that the assessment is measuring the same construct across populations?

- **Phase 6: Analyze Consequences**
  - Taken collectively, does the evidence gathered indicate that the assessment has achieved, for each population, the purpose or goals for which it was designed?
  - Taken collectively, does the evidence provide an understanding of unintended impact, whether positive, negative, or unknown?
  - How do interpersonal, interpersonnel, and ecological factors contribute to intended and unintended outcomes?

Note. The color figure can be viewed in the online version of this article at http://ila.onlinelibrary.wiley.com.
selected the case of the Diploma Examination in English Language Arts offered by the Ministry of Education in the Canadian province of Alberta, an assessment program that has been at the center of David’s program of research for the past 15 years.

As the ToA described in Figure 2 demonstrates, the Diploma Examination is embedded within a broader system of accountability developed by Alberta’s Ministry of Education. The overall goal of this accountability system is to promote excellence in educational leadership and teaching that leads to excellence in student achievement. Our purpose in articulating a ToA for this system of accountability is to make transparent the network of mechanisms, and the actions associated with them, that have been assumed to promote the Ministry’s envisioned outcomes. These mechanisms and actions can be systematically tested through the application of the IDAF heuristic.

Although the Government of Alberta has a well-articulated system of accountability, it has not published a ToA associated with that system. We begin our analysis, then, by developing a ToA informed by the work of Lewin, Corey, Argyris, and Schön, which we present in Figure 2.

The ToA for Alberta’s Diploma Examination program involves six program elements:
LITERACY ASSESSMENT FOR LEARNING

1. School-Level Accountability Pillar Results (A): Each year, Alberta Education ranks each school’s performance on 18 metrics that range from test performance, to graduation rates, to levels of parental engagement, to features of the school environment.

2. Individual-, Classroom-, and School-Level Score Reports (B): These annual reports document student achievement on standardized assessments. School-level reports are made available to the public.

3. Alberta Curriculum Documents (C): For each subject area, Alberta Education has developed a program of studies that explicates the learning objectives teachers must focus on in each subject area from grades 1 to 12.

4. Alberta Diploma Examination (D): Exams are designed to measure a sample of the curriculum outcomes in each core subject area.

5. Annual Diploma Examination Bulletins and Released Items (E): Each year exam managers for each diploma exam publish a report on the exam that provides an overview of the exam, trends in student performance, a test blueprint, scoring criteria, and sample test items.

6. Classroom Assessment Data (F): Teachers are expected to compile records of classroom assessment data that are then reported to Alberta Education. In courses that require a diploma exam, 70% of a student’s final grade is determined by performance on classroom assessments, and 30% is determined by diploma exam scores.

The hypothesized action mechanisms shown in Figure 2 (boxes G - R) describe the expected pattern of interactions between various stakeholders as they work to achieve the intended outcomes of Alberta’s Education system (boxes S–AG). The essential logic of the ToA is that various stakeholders take the information available to them through the program elements; make decisions (hypothesized action mechanisms) about resource allocation, program planning, and/or instructional focus; and identify anticipated negative and positive intermediate and long-term consequences.

Interrogating ToA Through an IDAF Lens

Whereas a ToA such as that shown in Figure 2 lays out the logic that takes us from program elements to intended policy outcomes, the IDAF shown in Figure 1 provides a mechanism for critically examining that logic. Integrating the models provides teachers with a tool kit to draw attention of assessment stakeholders to the components and consequences of assessment implementation. That is, the integrative model helps teachers do the following:

- Demand precision from the assessment that focuses on construct definitional clarity, construct curricular alignment, and construct curricular consequences
- Create heuristics for examining the technical quality of an assessment, its impact on populations, and the intrapersonal, interpersonal, and ecological factors that mitigate or enhance its impact

Given the space available in this column, in the sections that follow, we limit our examination to the sections of the ToA described in Figure 2 that pertain specifically to the role of teachers (boxes M, N, O, P, X, Y, and AC).

Investigating Program Elements

As shown in Phases 1 and 2 of Figure 1, central to IDAF is a focus on assessment aim and construct articulation. We emphasize this focus in the first stage of investigating the ToA because resonance between assessment aim and clearly articulated constructs provide the foundation for all program elements.

In Figure 3, we highlight the construct-focused questions that necessarily lead any investigation of a ToA. With respect to the writing component of Alberta’s English Language Arts Diploma Examination, Slomp (2008) found that the exam measured students’ ability to structure ideas, paragraphs, and sentences; their ability to develop and support ideas, to use diction, syntax, and punctuation for effect; and their ability to create semipolished text within tight time frames. He also found, however, that given the timed impromptu nature of the exam, it was not able to capture student ability to use exploratory writing or substantive revisions as key aspects of their writing processes. These aspects of construct underrepresentation make it difficult for teachers to draw meaningful inferences about student writing ability in non-testing contexts (Slomp, 2016).

We demonstrate in the following two sections how this construct underrepresentation in the Diploma Examination undermines the entire ToA and the achievement of the long-term consequences the assessment program was designed to achieve.

Investigating Hypothesized Action Mechanisms

Phases 3 through 5 of Figure 1, when applied to the ToA for the Alberta Diploma Examination in Figure 2, invite
teachers to raise questions about a given assessment design, its scoring system, and its results. A primary concern for teachers must be the alignment between the curriculum, the diploma exam, and teachers’ classroom assessment plans. Questions regarding this alignment are shown in Figure 4. Misalignment undermines capacity to achieve the intended outcomes described in the ToA. As one teacher observed in Slomp’s 2008 study, the exam narrowly focused on a student’s written responses to literature while ignoring other forms and
purposes for writing; it measured only students’ ability to write under pressure and severe time restrictions; and it involved writing under artificial conditions, with no access to resources and no opportunities for collaboration.

Relations between construct underrepresentation and curricular impoverishment are well known to literacy educators. What is less known are the ways, noted in the teacher’s observations, that construct underrepresentation can result in actual barriers to learning for groups of students. Returning to Figure 1, for example, teachers may decide to demand that test developers provide information on the performance of student subgroups (categorized by socioeconomic status, race/ethnicity, gender assignment, and other relevant groupings) to demonstrate that the exam does not create an occasion for disparate impact in requiring writing under constrained conditions (Poe & Cogan, 2016). Indeed, if a test developer fails to demonstrate that there are no subgroup mean differences, then there is a good reason—often legal in origin—that the test should not be used until follow-up studies are conducted to identify the causes of such differences (AERA, APA, & NCME, 2014). Construct underrepresentation can result in both a diminished curriculum and denial of opportunity to learn.

Figure 5
Examination of Intended Outcomes

Investigating Consequences
Finally, it’s important to ask questions regarding the intended policy consequences associated with an assessment program. As with the other phases of inquiry, IDAF prompts us to question the impact of the assessment program by focusing on questions related to Phase 6 of Figure 1. Figure 5 illustrates the types of questions that should be asked when investigating anticipated outcomes, and the matrix of factors that support or inhibit their achievement. Of special interest here are the intrapersonal, interpersonal, and ecological factors that shape teachers’ responses to the exam, that influence their engagement with professional development opportunities, and that shape their capacity to enact instructional and assessment change.

Examining the impact of Alberta’s Grade 12 academic English Diploma Examination from a system-level perspective, Slomp, Graves, and Broad (2014) found that teachers valued risk-taking—an intrapersonal factor associated with openness to writing experience—both in terms of their approaches to teaching and in terms of how they wanted their students to approach the act of writing. They also found that ecological pressures (assessment-based, often narrow
monitoring of schools and misaligned expectations of students, teachers, administrators, and parents) frequently prevented experimentation and risk-taking in their teaching of writing—two outcomes critically important to the creation of writerly identity (Forzani, Corrigan, & Slomp, 2020).

In examining assessment consequences, teachers may use our combined ToA and IDAF models to establish links among the technical quality of an assessment program, intrapersonal factors, and ecological factors. As this cases study illustrates, in many ways Alberta’s educational system works against itself. Put straightforwardly, the Diploma Examination itself undermines the ultimate discovery goals expressed in the ToA (Figure 2, box AG).

A systematic analysis of the ToA draws a line from assessment design to teacher actions, to student outcomes. In this case, that line places responsibility for negative unintended outcomes stemming from the assessment program with the assessment designers (where it properly belongs) rather than with teachers (where too often it is placed). When policymakers turn to testing as a way to improve student outcomes, teachers should react by asking those same leaders to show and explain their ToA figures.

Implications for Classroom Teachers
The late John Lewis was fond of saying that his work as a civil rights leader and U.S. representative was directed toward one end: to get in the way of trouble. To get there, you had to make good trouble. Anyone who has watched the two recent documentaries about his life (Dowdey, 2017; Porter, 2020) realizes that merely summoning our better angels will not get us anywhere. When they arrive, we will need to sit them down and explain to them how to work in activist traditions.

At this particular moment in history, we believe it is critically important to provide strategies for identifying assessment elements and interrogating assessment consequences. In the spirit of Lewin, Corey, Argyris, and Schön, we encourage educators to demand that ToA frameworks—such as those shown in Figures 3–5—accompany any imposed assessment. Unless assessment sponsors and developers identify a test’s components, vehicles, and consequences, score interpretation and use will be as likely to do as much harm as good. Furthermore, we encourage teachers to advocate for robust programs of research using the IDAF phases shown in Figure 1. Principled questions of inquiry regarding claims enliven ToA structures so that evidence can be gathered to inform—and qualify and, perhaps, challenge—the inferences test developers make about our students.

Recalling Lewis, we must find a way to get in the way of trouble and to make our schools a better place for all children. Now is the time to make good trouble.

REFERENCES


The department editor welcomes reader comments.

DAVID SLOMP is an associate professor in the Faculty of Education at the University of Lethbridge, AB, Canada; email david.slomp@uleth.ca.

---

**ILAS DIGITAL EVENTS**

Online Learning. Online Resources. Online Community.

- **Members Save More:** ILA members enjoy special rates and can access the ILA at Home Webinars for free!
- **Timely Topics:** Dig deep into the issues you’re facing right now in your classroom and district.
- **Interactive Discussions:** Participate in live Q&As, connect with other attendees, and have lively discussions in real time.
- **On-Demand Access:** Watch wherever you are, whenever you want. You can access on-demand recordings for up to a year.

**Featuring these literacy experts:**

Sara K. Ahmed, Nell K. Duke, Chad Everett, Steve Graham, Jimmy Kim, Donalyn Miller, Cornelius Minor, Ernest Morrell, Molly Ness, Timothy Rasinski, Kate Roberts, Timothy Shanahan, Nicole Patton Terry, Yong Zhao, and more!

Attend live or review on-demand recordings at your convenience!

literacyworldwide.org/DigitalEvents