Applying Design Thinking to systemic problems in educational assessment information management

W P Fisher Jr¹, Emily Pey-Tee Oon² and Spencer Benson³
¹ Research Associate, BEAR Center, Graduate School of Education, University of California, Berkeley, Berkeley California 94702 USA
² Faculty of Education, University of Macau, Taipa, Macau SAR, China
³ Centre for Teaching and Learning Enhancement, University of Macau, Taipa, Macau SAR, China

¹E-mail: wfisher@berkeley.edu

Abstract. The integrative process of design thinking provides a compelling framework for addressing problems in education involving coherent assessment and instruction. New possibilities for more meaningful developmental coherence within, horizontal coherence across, and vertical coherence between classroom formative and accountability assessments emerge by thinking through and enacting conceptual solutions that begin with the end in mind and that involve end users every step of the way. The nonlinear process involves empathy for students and teachers, the definition of a solvable problem (one with solutions that are feasible, viable, and desirable), the ideation of a brainstormed array of ideal “possibly impossible” resources for creativity (such as a very demanding measurement model and traceability to unit standards), the rapid development of an initial prototype expected to have limitations, and the first of an iterative series of tests that deliberately test the prototype en route to a recycling as needed through any other parts of the process. Applying Design Thinking may lead to new solutions to problems encountered in reconceiving and reconfiguring classrooms as meta-design ecosystem niches of creativity and innovation. An assessment outcome prototype drawn from recent work with teacher/educators that integrates developmental, horizontal, and vertical coherence is presented.

1. Introduction
A broad-based platform for creativity termed Design Thinking [DT] combines aspects of art, engineering, and empathy with persons experiencing systemic problems possibly amenable to new solutions. DT is associated with IDEO and the Hasso-Plattner Institute of Design at Stanford University. Some advocates suggest DT may become so widely adopted it becomes a “New Liberal Arts” [1]. Others insist its fusion of liberal arts and professional education mean DT will come into full flower only at the systems level, across the whole university, breaking down artificial divisions (silos) [2]. This focus on enhanced information flows and communications across horizontal levels has much in common with efforts aimed at increasing the coherence of educational outcome assessments [3-4] and with the ecosystem conception of organizational and interdisciplinary interdependencies and interrelations [5-8].

Toward this end, the five aspects of value creation integrated in DT; “empathy-define-ideate-prototype-test,” were employed to devise new approaches to problems associated with coherence across
three primary educational assessment use cases. Developmental coherence involves individual students’ growth along learning progressions; horizontal coherence concerns the comparability of learning progressions and outcomes across students, classrooms, and schools; and vertical coherence aligns formative classroom and accountability assessments at regional, national, and international levels [3,4]. Accountability concerns test uses at the individual, class, school and state levels to monitor and assess the degree to which students are meeting learning outcomes and being taught by effective teachers. Though these three domains of assessment cannot be combined by means of a universal test, their common focus on educational processes and outcomes suggests the viability of greater degrees of coherent alignment of available information with the management and policy needs of individual students, classrooms, schools, districts, etc. After all, mathematics ability is referred to as a single thing across curricula, students, tests, schools, even though it is represented incoherently in different terms across and within all of these contexts.

The multilevel nature of meaning across local, midrange, and broader contexts has long been of philosophical interest [5,6]. Communication systems are ecologies in which micro-level processes within individuals are distinct from, but interact with, meso-level processes between individuals, which in turn are distinct from but interact with macro-level group processes. Developmental, horizontal, and vertical coherence issues in educational assessment correspond to these varying levels of complexity in communication systems. Practical applications of these distinctions in epidemiology [7] are leading to new, productive relationships between clinical medicine and public health efforts [8]. The primary justification for taking an ecological approach to problems of coherence in educational assessment is to understand learning in the contexts of its various environments. Contexts affect learning in a variety of ways that cannot be grasped from measures of individual attributes. Broad-scale efforts at sorting out the sources of distinct classes of effects on learning can only begin when the appropriate communications systems are in place.

In the DT context, the viability of more coherent educational assessment information must be complemented by the technical feasibility of the solutions and the desirability end users have for them. The technical challenges are complex, as they involve not only the comparability of different assessments administered to different students, but also comparability across different formative and accountability assessment purposes. Feasibility is supported, however, by a range of available technical solutions, including adaptive administration of item banks [9], theory-driven automatic item generation and inference in the context of metrological traceability [10,11], and test equating methods encompassing a variety of item formats [12].

The desirability of viable and feasible design solutions depends to a large extent on empathy with end users. Empathic identification does not always involve taking end user opinions literally, however. As Henry Ford is reputed to have observed, if he had listened to what people said they wanted in the way of improved transportation, he would have focused on faster horses. It accordingly happens that virtually every major technical innovation, from automobiles to electricity to the telegraph, telephone, television, and the Internet, has emerged in a context in which there was no public demand whatsoever for a line of products that later came to seem indispensable. Can this kind of result can be obtained via the application of DT to the development of new technologies for educational information applications?

2. The DT integrative process

DT is not a linear sequence of steps. One might take up empathy, definition, ideation, prototyping, and testing in almost any order, with variation in emphasis across them, both within and across iterations. The educational example described here moves through an artificial sequence solely for the purpose of orientation to the different aspects of DT integrated into a process.

2.1 Empathy

A number of conceptual, organizational, and technical barriers result in teachers being overburdened with repetitive tasks that follow largely from the need to micromanage the details of every student’s
learning, and of every assessment of that learning. Compared with the information available from well-made formative assessments integrated with learning materials [13], students and teachers alike find it difficult to know what lessons are most appropriate for which students.

The quality of information on educational processes and outcomes that is available in principle is vastly superior to the quality of information actually used by teachers. Low quality information makes it difficult to match readers to texts, but it makes plotting development, comparing results with other teachers, and setting achievable accountability goals virtually impossible. Teachers are being held accountable for managing outcomes they cannot effectively visualize and compare day to day, either within their own classes, or across classrooms. This situation is an example of what Bateson [5,6] called the transcontextual syndrome, a double bind imposed by the figurative schizophrenia of demanding, for instance, a behaviour coherent within a horizontal frame of reference that is simultaneously incoherent in the developmental frame of reference.

2.2 Define
The problem found in empathizing with teachers then concerns the incoherence of the available learning outcome management information systems. What would a coherent education management information system look like? Developmental, horizontal, and vertical coherence are needed to address the learning processes of individual students, to enable comparability over time and across classrooms, and to manage for long term accountability goals. Coherence of this kind demands meaningful and comparable quantitative measures substantively interpretable in terms of learning progressions. Temporarily setting aside concerns for viability, feasibility, and desirability, what kinds of resources might be useful in formulating solutions to these problems of coherence?

2.3 Ideate
Starting from the beginning, how do these coherence problems fit into the larger context and goals of education? Education is supposed to provide students with experiences that enable them to successfully address problems in real life, to create value for themselves and others. No one thinks students need to encounter every possible life problem in the classroom. A curriculum samples from the infinite universe of all possible problems and lessons in ways intended to represent real life well enough to prepare students for adult responsibilities.

Conversely, schools are supposed to produce consistent learning outcomes independent of the particular students in attendance. The usually unstated but widely adopted assumptions are, then, that (a) education takes place over time as a process requiring the conjoint independence of student abilities and the difficulties of the lessons to be learned, such that the probability of student success is conditional on the difference between ability and difficulty, and (b) that this independence is never perfectly realized, however useful it has proven historically as a heuristic fiction.

In this context, how might the problem of educational information coherence be formulated? What not widely considered, “possibly impossible” solutions might be brainstormed as resources for later trial and error applications? Where might audacious models for proceeding be found? Are there approaches in science, engineering and DT that have previously been applied in situations like this, where problems and solutions, questions and answers, exist in similar kinds of mutual, interdependent relations?

Qualitatively speaking, the situation is analogous to the general form of a wide range of measurement models, including, for instance, Newton’s laws [14]. Psychometric models in this form have lately been shown to share important features with models and principles widely used in engineering and science [15,16]. Could the design problem of coherently related educational assessments be successfully addressed in terms of media formatted in a common language and distributed throughout a communications network? Historical and social studies of science have led to an emphasis on the capacity of researchers to coordinate and align their behaviours and decisions virtually, by means of common languages and shared values that do not homogenize but incorporate local variations [17,18]. Extending that work, this kind of ecological networking has been proposed as an alternative approach
to addressing contemporary issues in psychology and the social sciences [19], suggesting efforts aimed at devising ecosystems of educational forms of life inhabiting diverse niches may be worth pursuing.

Each form of coherence, and different approaches to configuring their expression, will vary in its parameterization. A model that explicitly states the joint relationship of student abilities and item difficulties, and makes the potential portability of that relationship testable, takes the form

$$\ln[P_{nj}/(P_{n;i})] = B_n - D_j$$

which says that the log-odds of observed success for student n on item i at partial credit score j is equal to the difference between the estimate B of person n's ability and the difficulty estimate D of item i at rating j [20,21]. In this kind of model, the group-level construct emerges as a self-organized pattern from the within-individual responses. More complex hierarchical models of multilevel systems are available [22]. A key point is that, in the same way that no drawn triangle conforms perfectly to the Pythagorean theorem, no data ever fit a model of this kind. The question is whether the model is useful.

The solution to the problem of coherence in educational assessment hinges on identifying what is already known about student learning, encapsulating it in a portable technology, and making the information embodied in this device useful in different forms across the developmental, horizontal, and vertical contexts for instructional, quality improvement, and accountability applications. Prototyping is the process in which technical solutions of this kind are proposed.

2.4 Prototype

For students, the problem of knowing what to study next is one of locating themselves on a learning progression and checking for special strengths and weaknesses in need of attention. For teachers, embodying learning progressions on portable tools gauged in common metrics makes it possible to more easily relate student abilities to appropriate learning materials. Figure 1 shows a self-scoring form [23] produced from widely used educational measurement data analysis software. Though these kinds of forms have a long history of use in education [24], they are not widely known. An automated application could work from computerized test data to indicate on this form which questions a student answered correctly and incorrectly, as well as showing the associated correct answer count, measure, uncertainty,

| NUM | ITEM CONTENT SUMMARY |
|-----|----------------------|
| 0   | 100* MOST DIFFICULT ITEM HERE |
| 0   | 101*                     |
| 0   | INCREASING DIFFICULTY   |
| 0   | 111*                     |
| 0   | 110*                     |
| 0   | 107*                     |
| 0   | 104*                     |
| 0   | 101*                     |
| 0   | 99* LEAST DIFFICULT ITEM HERE |

| MEASUREMENT SCALE |
|-------------------|
| COUNT CORRECT |
| UNCERTAINTY |
| STUDENT |
| MEASURE |
| DISTRIBUTION |
| MEAN (M), 1 SD (S), 2 SD (T) |
| PERCENTILE |

Figure 1. Prototype individual student self-scoring form for developmental coherence.
Figure 2. Prototype classroom comparison report for horizontal coherence.

![Horizontal Coherence Report](image)

Figure 3. Prototype multilevel comparison report for vertical coherence.

![Vertical Coherence Report](image)

and percentile rank relative to an overall distribution. Advanced users may want to include data consistency indices (model fit statistics), and other diagnostic, sample, and population statistics. It is important to note here that, in accord with results observed in studies of universals in medicine [18] and other fields, individual student response patterns will rarely conform strictly to the expected pattern.

These applications in the domain of developmental coherence at the micro level within individual students require levels of detail and precision in the information reported not shared by the demands for horizontal and vertical coherence at the meso and macro levels. For horizontally coherent classroom management, teachers need actionable information identifying opportunities for improvement at the student level, as well as from comparisons with other teachers’ results (Figure 2). Plots of individual student performance over time, and cross sectional plots of multiple students at a given point in time (as shown in the student measure distribution in Figure 1), will support meaningful comparisons.

Finally, as shown in Figure 3, for vertically coherent management up and down a hierarchy of accountability demands, everyone involved should be able, if needed, to drill down from any given level
to the patterns of expected performance on the learning hierarchy for any subgroup of students, or any individual student. The vertical reach of the report could be extended to statewide, national, and international level. Interpretive guidelines will include construct maps and instructions for use.

2.5 Test
Prototype tests are proceeding with focus sessions on the to-date prototype developments. Trial runs with real data will be shown to teachers, who will evaluate the strengths, weaknesses, opportunities, and threats that may conceivably be expected to emerge relative to the proposed tools. A second pass will then be performed using data accumulated by a single teacher over time as a demonstration of the value obtained relative to the pre-existing methods and tools. The results from this second test will be fed into a new focus session in a third pass, and revisions from that test will be used in a pilot deployment.

3. Education ecosystem implications
The coherent coordination of information across levels and silos in social ecosystems implies a view of measured constructs as complex adaptive self-organizing forms of life. Identifying, documenting, and embodying these species’ and niches’ interdependencies could lead to important new opportunities for the improvement of educational outcomes.

References
[1] Miller P N 2015 Chronicle Higher Educ 61 29
[2] Spinelli S Jr and Nixon N 2015 Chronicle Higher Educ 61 31
[3] Gorin J and Mislevy R 2013 Inherent measurement challenges in the next generation science standards for both formative and summative assessment (Princeton, NJ: ETS)
[4] Wilson M 2004 Towards coherence between classroom assessment and accountability (Chicago: University of Chicago Press)
[5] Star S L and Ruhleder K 1996 Information Systems Research 7 111-134
[6] Bateson G 1972 Steps to an ecology of mind (Chicago: University of Chicago Press)
[7] Susser M and Susser E 1996 Am J Public Health 86 674-677
[8] Bizouarn P 2016 Médecine/Sciences 32 500-505
[9] Barney M and Fisher W P Jr 2016 Ann Rev Org Psychol Org Behav 3 469-490
[10] Stenner A J, Fisher W P Jr, Stone M and Burdick D 2013 Frontiers in Psychol: Quant Psychol Measurement 4 1-14
[11] Dawson T 2004 J Adult Dev 11 71-85
[12] Öztürk-Gübes N and Kelecioglu H 2016 Educational Sciences: Theory & Practice 16 715-734
[13] Griffin P 2007 Studies Educ Eval 33 87-99
[14] Fisher W P Jr 2010 J Physics Conf Series 238
[15] Mari L and Wilson M 2014 Measurement 51 315-327
[16] Pendrill L and Fisher W P Jr 2015 Measurement 71 46-55
[17] Latour B 1998 Remaking reality: Nature at the millennium ed N Castree and B Willems-Braun (New York: Routledge) pp 221-242
[18] Berg M and Timmermans S 2000 Configurations 8 31-61
[19] Fisher W P Jr and Wilson M 2015 Pensamiento Educativo 52 55-78
[20] Rasch G 1980 Probabilistic models for some intelligence and attainment tests (Chicago: University of Chicago Press)
[21] Andrich D 1978 Psychometrika 43 561-573
[22] Beretvas N and Kamata A 2007 Rasch measurement: Advanced and specialized applications ed E V Smith Jr and R M Smith (Maple Grove, MN: JAM Press) pp 290-470
[23] Linacre J M 1997 Physical Medicine and Rehabilitation State of the Art Reviews 11 315-324
[24] Connolly A J, Nachtman W and Pritchett E M 1971 Keymath: Diagnostic Arithmetic Test (Circle Pines, Minnesota: American Guidance Service)