Conceptual frameworks for reporting results of assessment activities.

Paper Presented at the Annual Meeting of the National Council for Measurement in Education

John T. Behrens – Pearson (John.Behrens@Pearson.com)

Kristen DiCerbo – Pearson (Kristen.DiCerbo@Pearson.com)

Dan Murphy – Pearson (Dan.Murphy@Pearson.com)

Dan Robinson – Colorado State University (Dan.Robinson@colostate.edu)

April 2013

ALWAYS LEARNING

PEARSON

About Pearson
Everything we do at Pearson grows out of a clear mission: to help people make progress in their lives through personalized and connected learning solutions that are accessible, affordable, and that achieve results, focusing on college-and-career readiness, digital learning, educator effectiveness, and research for innovation and efficacy. Through our experience and expertise, investment in pioneering technologies, and promotion of collaboration throughout the education landscape, we continue to set the standard for leadership in education. For more information about Pearson, visit http://www.pearson.com/.

About Pearson's Research & Innovation Network
Our network **mission** is to spark innovation and create, connect, and communicate research and development that drives more effective learning. Our **vision** is students and educators learning in new ways so they can progress faster in a digital world. Pearson's research papers share our experts' perspectives with educators, researchers, policy makers and other stakeholders. Pearson's research publications may be obtained at: http://researchnetwork.pearson.com/.

Abstract

In this paper we discuss three lenses for understanding the interaction of the goals of assessment developers and assessment stakeholders: Information communication, social activity, and educational literacy. Information communication frames the designer-stakeholder interaction as a knowledge transfer problem that focuses on transmitting non-distorted messages from the sender to the receiver. There has been growing interest and accumulated recommendation in this area (Hambleton, 2002, Hambleton & Slater, 1997; Jaeger, 1998; Vezzu, VanWinkle, & Zapata-Rivera, 2012) and we propose some language to consider the various recommendations. The social activity lens recommends activity theory as a theoretical framework with which to approach the design of assessment systems. This approach takes into account a number of social roles and epistemic frames that individuals and groups may bring to the interaction with the assessment system.   Broadening the understanding of the social role of assessment further, in the final section we suggest the framework of Health Literacy as a possible model for considering the role of assessment activity and assessment information in a larger context of Educational Literacy.

Keywords: assessment reporting, literacy, communication

Introduction

As the field of assessment advances and matures, the breadth of relevant inputs, scope, and concerns expands. New types of conceptualizations of the assessment endeavor have evolved (Mislevy, Almond, & Steinberg, 2003) as well as new extensions of focus into areas such as understanding the role of technology in assessment practice and approaching new constructs such as Twenty-First Century Skills (Mayrath, Clarke-Midura, Robinson, & Schraw, 2012). Of particular interest is the notion of taking responsibility for the communication of assessment results to the receiving stakeholder in the form of score reports or other communication systems (Zenisky & Hambleton, 2012).

In this paper we conceptualize the reporting of assessment results as occurring in three conceptual layers. The first layer borrows from notions of evidentiary argument and reasoning combined with notions of individual functioning and interpretation of information from cognitive psychology. In this approach, the motives and engagement of the user are more or less assumed and the goal is to create interactions that communicate the information to the receiver in a manner that minimizes distortion of the information transferred. We call this the Information Communication layer.

A second set of concerns arises when the interpretive frames that individuals and groups bring to the assessment system may vary or are not well understood. Borrowing from the literature on human computer interaction we recommend concepts based on Activity Theory (Engström, Miettinen, & Punamaki, 2007) and illustrate their use by way of Activity System diagrams. Because this layer of design and efficacy concerns bringing together sociological and anthropologic concepts, we call this the Social Activity layer.

Finally, a third layer of relevant concerns may be raised that consider the role of assessment systems and assessment communications in the larger context of fostering educational change in general. This layer considers how assessment information is

expected to interact with other types of educational information and the role the end user is expected to play in the educational system as a whole. This layer is conceived in terms of the analogous concept of Health Literacy as applied to an idea of Educational literacy and is referred to as the Societal Transformation layer.

Primary Constructs	Core Question	Name
Logic / Comprehension	What do designers want to communicate?	Information Communication Layer
Activity System	How are people approaching and interpreting the system?	Social Activity Layer
Educational Literacy	What is the desired effect in the educational system?	Societal Transformational Layer

In the following sections we discuss each of these conceptual frames for understanding and designing assessment reporting systems and highlight implications of using the various frames.

Information Communication

To date, the most common method for conceptualizing assessment reporting is from the perspective of communicating results of assessment activities as understood by the designer of the assessments. In this view the primary goals of the reporting effort are to (1) clarify goals (Bradshaw & Wheater, 2009), (2) align communication elements with assessment goals (Goodman & Hambleton, 2004; Hambleton & Slater, 1997; NEGP, 1998; Wainer, Hambleton, & Meara, 1999), (3) minimize error in the conceptualization, computation or explanation of results (Ryan, 2006), (4) avoid ambiguities in language or display that would lead to misinterpretation (Goodman & Hambleton, 2004; Hambleton & Slater, 1997; NEGP, 1998; Wainer, et al., 1999), (5) warn against misuse (Ryan, 2006), (6)

consider individual variations in necessary skills such as language and numeric literacy (Hambleton & Slater, 1997; Jaeger, 2003; Trout & Hyde, 2006; Zenisky, Hambleton, & Sireci, 2009; Underwood, Zapata-Rivera, & VanWinkle; 2007, Vezzu, et al., 2012) and (7) use an empirical process to verify appropriate receipt of the message (Hambleton & Slater, 1997; Trout & Hyde, 2006; Wainer, et al., 1999).

These and other recommendations or guidelines for assessment results reporting can be organized into three main categories: Interpretive Constraints, Communicative Considerations, and Process Imperatives. Figure 1 depicts the conceptual organization of the Information Communication perspective.
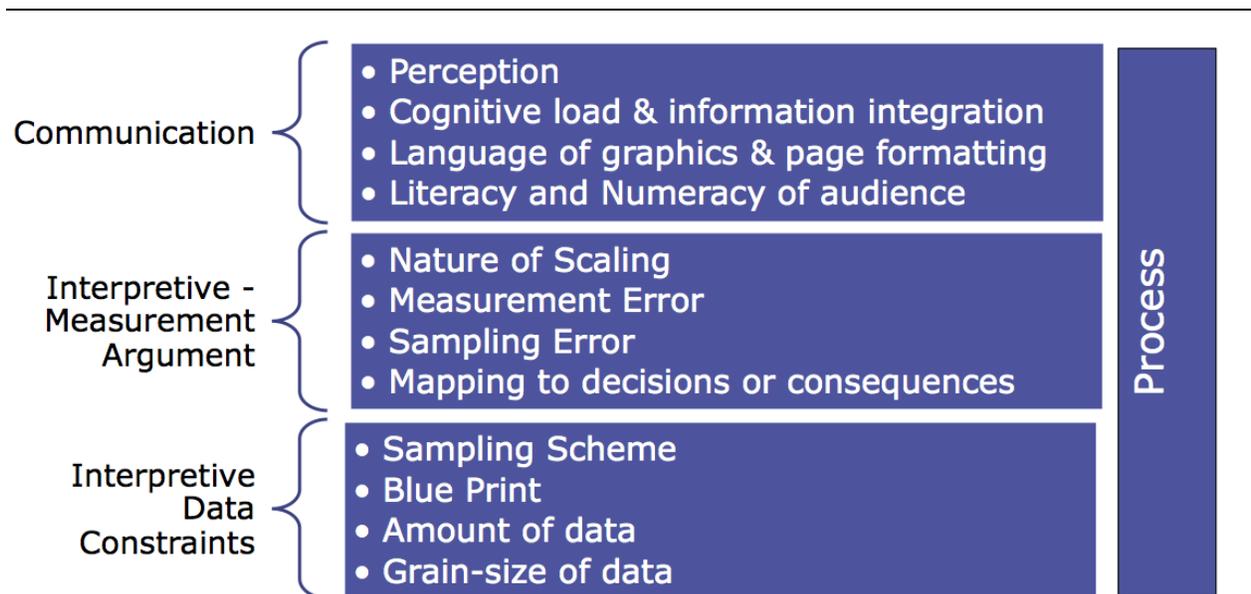

Figure 1. Elements of the Information Communication Approach.

Interpretive Constraints

Considering assessment as an evidentiary argument (Mislevy, 1994) opens the question regarding what are the relevant inputs to the argument and what are the range of appropriate inferences that may follow. Here, the many variations in the goals, uses, and mechanics of assessment need to be considered. In general, the interpretive constraints can be considered at two stages as well: interpretive constraints based on the data at hand

and interpretive constraints based on the logical-measurement arguments made from that data.

We think that it is important to separate these two issues as the measurement aspects of the assessment are dependent on the data at hand but often multiple assessment argument systems may be used. For example a specific set of data may be modeled variously using Item Responses Theory (Lord, 1980), Classical Test Theory, or other computational approaches (Mislevy, Steinberg, & Almond, 2002).

Issues related to the data that serve as inputs largely fall under the heading of task models (Mislevy, Steinberg, & Almond, 2003). These interpretive constraints include a broad range of issues including (a) understanding the source and veracity of the data, (b) understanding the form and interpretability of the data, (c) understanding data granularity in order to align it with subsequent inferences, and (d) amount of data associated with individuals or individual tasks either in aggregate or in terms of specific sampling schemes. Some sampling schemes may be fixed (e.g., a fixed form test), whereas others may be variable based on a specified optimization algorithm (e.g., Computer Adaptive Testing approach).

Issues related to measurement argument are associated with the elements of the Conceptual Assessment Framework (Mislevy, et al., 2003) which relates the features of the data stream to observable variables which, in turn, are related to latent user- (or student- or learner-) model variables. While these elements and the related processes typically result in scale scores, other transformations or projections may lead to representation such as probability of differential category membership (Mislevy, 1994).

Although to the trained assessment designer or psychometrician, the enumeration of these issues as interpretive constraints may be considered obvious and clearly stated as "only say what is justified", we believe it is essential to make explicit these foundational processes in this conversation because new technologies for assessment delivery and results communication support the easy transfer of messages that may not conform to

these understandings.   The continued lowering of computer system development costs means it is increasingly easy to establish a system that reports results without considering or understanding the interpretive measurement issues that should undergird the communication. It is important for all assessment designers to remember that regardless of the complexity of the claims they are making regarding learner knowledge, skills, and attributes (Behrens & DiCerbo, 2013) the claims rest on a chain of logic building off of these fundamental interpretive constraints.

Communicative Considerations

Having established the range of appropriate interpretations of data and data transformations, communication regarding the outputs of the assessment process will typically move to a series of considerations regarding the appropriate receipt and interpretation of the message. In this area of concern, the assessment community has expanded its understanding and practice to include a number of cognitive issues such as information display, numeracy, and text literacy. Researchers such as Larkin and Simon (1987), Shah, Mayer, and Hegarty (1999), and Winn (1988) have examined the perceptual and cognitive advantages of various display types in communication. Somewhat paradoxically, more efficient and effective forms of communication can sometimes lead to poorer long-term understanding (e.g., Robinson & Schraw, 1994).

There is considerable evidence that receiving stakeholders have difficulty interpreting assessment reports. It has been found, for example, that too many technical terms and concepts can confuse and intimidate report users who do not have the appropriate prior domain knowledge (Goodman & Hambleton, 2004; Hambleton & Slater, 1997; NEGP 1998). On the other hand, there also may be a trade-off between reducing cognitive load and increasing engagement with the data. Appropriate complexity may encourage the stakeholder to process the information more deeply (Zapata-Rivera, & Zwick, 2011).

Process Imperatives

A third area of concern and advancement in the understanding of assessment reporting as a communicative function is the realization of the need for empirical validation of pre-supposed communicative functioning. For example, Goodman & Hambleton (2004) argued that reports should be piloted with members of the intended audience and Hattie (2009) argued that "evidence is needed to demonstrate how readers are interpreting reports".   Patelis & Matos-Elefonte (2009) discussed a long process of empirical validation of the test construction and corollary communicative extensions.

Social Activity

The assessment-based understanding of results reporting described above starts with the conceptualization of the assessment designers and users from the perspective of intended goals and uses, primarily from the perspective of measurement and related disciplines. The Social Activity lens borrows from the Activity Theory literature (Engström, Miettinen, & Punamaki, 2007) in an attempt to broaden the range of inputs and relevant considerations. In this section we discuss some of the fundamental constructs of Activity Theory and consider a number of possible implications for assessment reporting that may flow from the application of this lens.

Activity Theory suggests an activity (a motivated or goal based endeavor) as the fundamental unit of social analysis in opposition to traditions that might seek to understand human action in isolation. Writing in the instructional literature, Jonassen, Tessmer, and Hannum (1999) note:

> Activity cannot be understood or analyzed outside the context
> in which it occurs. So when analyzing human activity, we must
> examine not only the kinds of activities that people engage in,
> but also who is engaging in that activity, what their goals and
> intentions are, what objects or products result from the activity,

the rules and norms that circumscribe that activity, and the

larger community in which the activity occurs.

Jonassen et al. (1999, p. 160)

Accordingly, this framing provides a broader social lens than the elements described in the previous section which assumed a perfect alignment between funder, designer, and user goals and perspectives.

Activity Theory Overview

Activity Theory is a theoretic framework that centers on the idea of the interaction between a person or group (subject), a goal, motivation, or problem (object) and mediational interaction with tools, as indicated in the Subject ←→ Object ←→ Tool triangle at the top of Figure 2. A key feature of this approach is the consideration of the interactional features of the concept: Namely, not simply that an interaction occurs between the three aspects of the Activity System, but that the interaction reciprocally affects each of them as well. The subject's understanding of the object may change as they interact with different tools. Conversely, a subject's understanding of their role in the system may change as a function of the interaction with tools and objects.

The lower portion of Figure 2 highlights the social-interactional aspects of Activity Systems. In this portion, reference is made to the social norms, relevant communities, and divisions of labor. Again, a transactional mediated view is that each of these not only provides fixed context in which activity is situated, but also provides interaction and mediation of self-understanding.

Application

Application of the Activity System view typically occurs as an epistemic lens that is brought to bear on the interactional aspects of human activity. Frezzo, Behrens, and

Mislevy (2009) used the AT lens to identify various actors, objects, and tools in an instructional and assessment ecosystem that relied heavily on the use of a complex micro-world simulator. Application of this model allowed them to identify the variation in goals that different "users" held and how variation in interpretive frames led to some disconnects between the design characteristics of tools and the uses individuals put them to. For example, these authors discuss how teacher conceptualizations of the "rules" for formative assessments may be over-generalized to the application of those rules to summative situations in which a different set of rules were expected by other agents (subjects) in the activity system. This rationale led to the evolution of the system to include additional formative feedback in the assessment loop as well as additional clarification of top-down rules and their justification.

This approach of understanding human interaction with systems does not start with idealized notions of the perfect system following some platonic ideal, but rather understands that the ultimate use of the system will occur in the context of the self-understanding of the subjects which is affected by local (perhaps professional) norms, reference communities, and roles (reflecting division of labor). Individuals may adopt one set of cultural norms and professional identity at a primary job but use different norms or identify with different professional communities. The successful deployment of systems depends on understanding these variations in understanding identity, agency, and normative and extra-ordinary behavior.
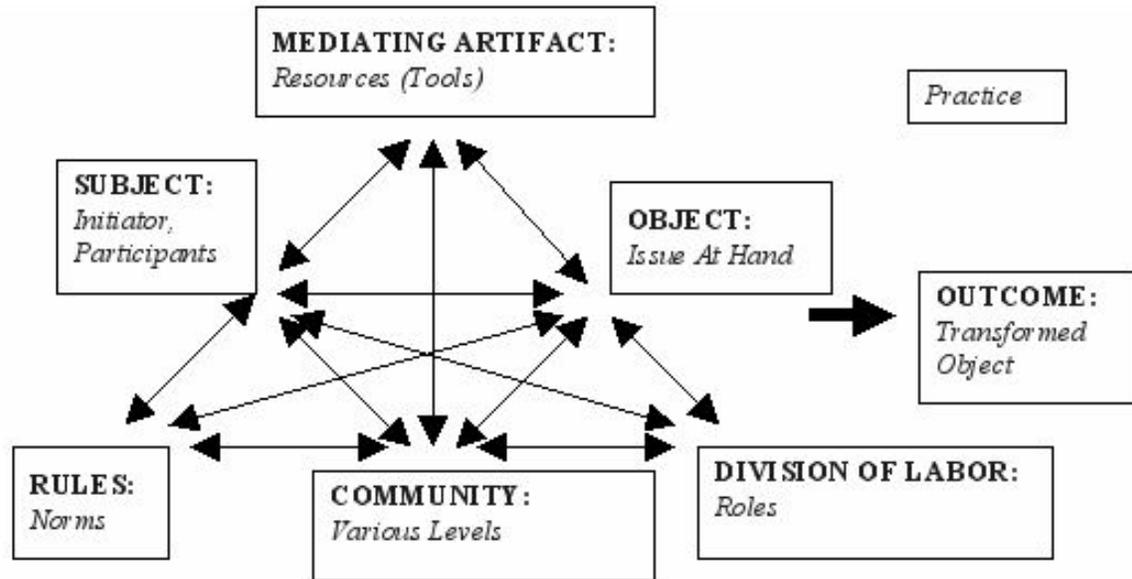
Figure 2. Activity System Diagram following Engeström (2007).

Frezzo, et al. (2009) also discussed the notion of the intersection of different activity systems in large system deployment. As illustrated in Figure 3, these authors discuss the idea that different actors in the system come with their different interpretations of system use and purposes and that successful system implementation from the perspective of one layer of the meta-system often depends on understanding the variations in belief and behavior in the other. If system designers understand the purpose and affordances of a system in a manner different from that of the users, the system may be judged as successful by some communities but unsuccessful by others.

This is a common occurrence in the deployment of assessment systems in which policy-level stakeholders are provided rigorous services that align to their broad goals related to accountability, while teachers and families may view a system as a failure given its lack of alignment with short-cycle instructional practice. The conceptual model of Activity Theory and the Activity System diagram provide a framework to understand the

necessary dimension of social meaning and activity that needs to be considered in designing
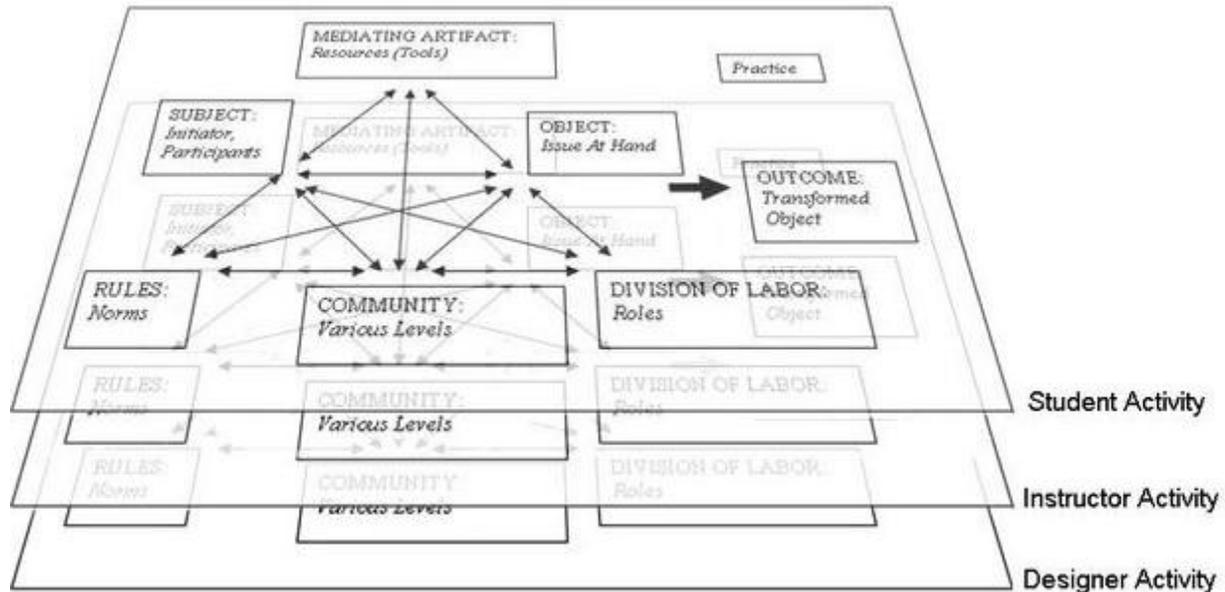
systems for human use.

Figure 3: Stacked Activity System Diagrams from Frezzo et al. (2009).

Implications

A number of possible implications for assessment communication may follow from

this analysis. First, we believe it is helpful that this view provides an auxiliary conceptual

frame to the common person-object framing common in many forms of psychology.   This

view emphasizes the purposeful nature of the individual in the system that recognizes that

individual goals and perspectives may not always align with goals and perspectives of either

designers or funders. One implication of this view is that when designing systems, it may

be important to understand the user's motivations and assumptions to not only

communicate about what the system is intended to do and how to use the system, but also

to communicate what range of likely erroneous assumptions or interpretations should not be

brought to the interaction. For example, summative assessment systems are often approached as if they had formative purposes, leading to possible confusion or frustration. The Activity Theory framing suggests not only that the assessment designer should communicate the constraints of appropriate interpretation given the assessment design, but that in many cases the work may be more effective if the underlying (possibly erroneous) beliefs of the end user were considered.

Second, this approach enumerates a series of dimensions and their interactions that should be considered in understanding possible forms of unintended interpretation or use, thereby holding the possibility that interaction design can be improved by understanding the frames from which the end users approach the interactions. This broadens the consideration beyond data use but also allows for the discussion of differential epistemic frames and ranges of plausible interpretations. Consider for example the notion of roles in the Activity System diagram. While a number of authors have argued for the importance of aligning system requirements with clear social roles such as parent, teacher, or administrator (e.g., Underwood, Zapata-Rivera, & VanWinkle, 2010), the Activity System approach asks whether there are additional social roles that these types of individuals may also undertake as they approach the system. For example, does the teacher approach the assessment system from the viewpoint of an advocate for the student, an advocate for the parent, and advocate for themselves, and if so how would each of these different socio-political perspectives change their expectation of information and what mental models of data analysis and presentation would correspondingly be called upon?

Third, we suggest the idea of starting with the range of end user systems, rather than the range of designer-created systems opens the possibility to consider new conceptualizations. Consider, for example the notion of the assessment system as an accountability monitoring system versus an assessment system as an ongoing learning support system. The first conceptualization is likely to emphasize broad sampling of the

content domain at specific points in time with weak linkages to specific actions in the classroom. The second conceptualization may need to emphasize ongoing data collection and student model updating with more idiosyncratic and likely less generalizable conclusions. In each of these cases the idea of decision support may vary widely as the systems would have been designed to support different types of decisions.

Of particular interest is the utility of certain forms of interaction and communication. For example, the results of summative tests are frequently communicated to parents in terms of scaled-scores which represent a complex transformation from data to numeric form. It is unclear what the theory of action is that suggests this is a format aligned with the goals of this community. Scaled scores have excellent measurement properties that allow for the compact and portable meaning regarding performance, but typically need transformation back into proficiency language for many stakeholders. An activity centric understanding of the reporting interaction would question whether that is the appropriate metric to convey at all.

The tool-centric and interactional aspects of the Activity System model also suggest the possibility of a number of reconceptualizations. Consider for example the common language regarding the communication of assessment results as "score reporting". This language places the score at the center of the value proposition and the inter-relationship between the user and the system. An alternate conceptualization might be communication as applied business intelligence or decision support. A decision support approach requires a clear understanding of the types of decisions that are relevant to the system and that would be explicitly supported. At present, most assessment systems provide decision support in only the most distal manner by providing baseline inputs (such as scores) to decision makers (or consultants) who create their own decisions in a variety of different ways using a variety of data inputs. A more direct decision approach would specify what types of

decisions are being supported, what types of data are most relevant, and provide tools for the analysis and integration of information directly relevant to those decisions.

Educational Literacy

In the first section, we discussed inputs to the assessment argument and communication given agreement on goals and motivations and conceptual frames. In the second section we discussed a framework to consider variation in goals and interpretations and to consider the social context of system interaction. The Information Communication lens could be considered to address the question: Assuming common conceptualization and motivation, what should be considered in communicating results? The Social Activity lens addresses the question: "Assuming variation in stakeholders across social groups, what are some of the key dimensions to consider in system conceptualization and design"? In this section, we broaden the question even further and ask "What is the social transformation we are expecting from the interaction"?

In many assessment situations, the argument for use of an assessment is implicitly assumed and rarely explicitly communicated. This will often lead to a lack of articulation of the theory of action driving the understanding and actions of the role of the end user: Why are we creating the assessment? Why are we reporting it in the manner we are? How do we understand these tools to be enabling or disabling the stakeholders in their progression toward their goals? To understand whether our assessment and communication achieves its desired effect we need to hold a theory of the desired effect which articulates desired transformations of the end user so we can measure our move toward that. Accordingly we want a model of user action and change that relates the inputs of the assessment interaction with the end goals of the educational endeavor overall.

To provide a language for considering the role of the assessment communication in the larger context of desired educational transformation of the end user, we recommend considering the concept of Health Literacy found in the health literature following such

authors as Nutbeam (2008). Figure 4 is a diagram from Nutbeam (2008) that describes

what he calls an "Asset Model" of Health Literacy. In this diagram we see a relatively

comprehensive overview of both the types of inputs that go into this model, as well as a

specification of the desired impacts, both proximal (tailored communication) and distal

(improved choices and outcomes) as well as a specification of moderator variables (changes
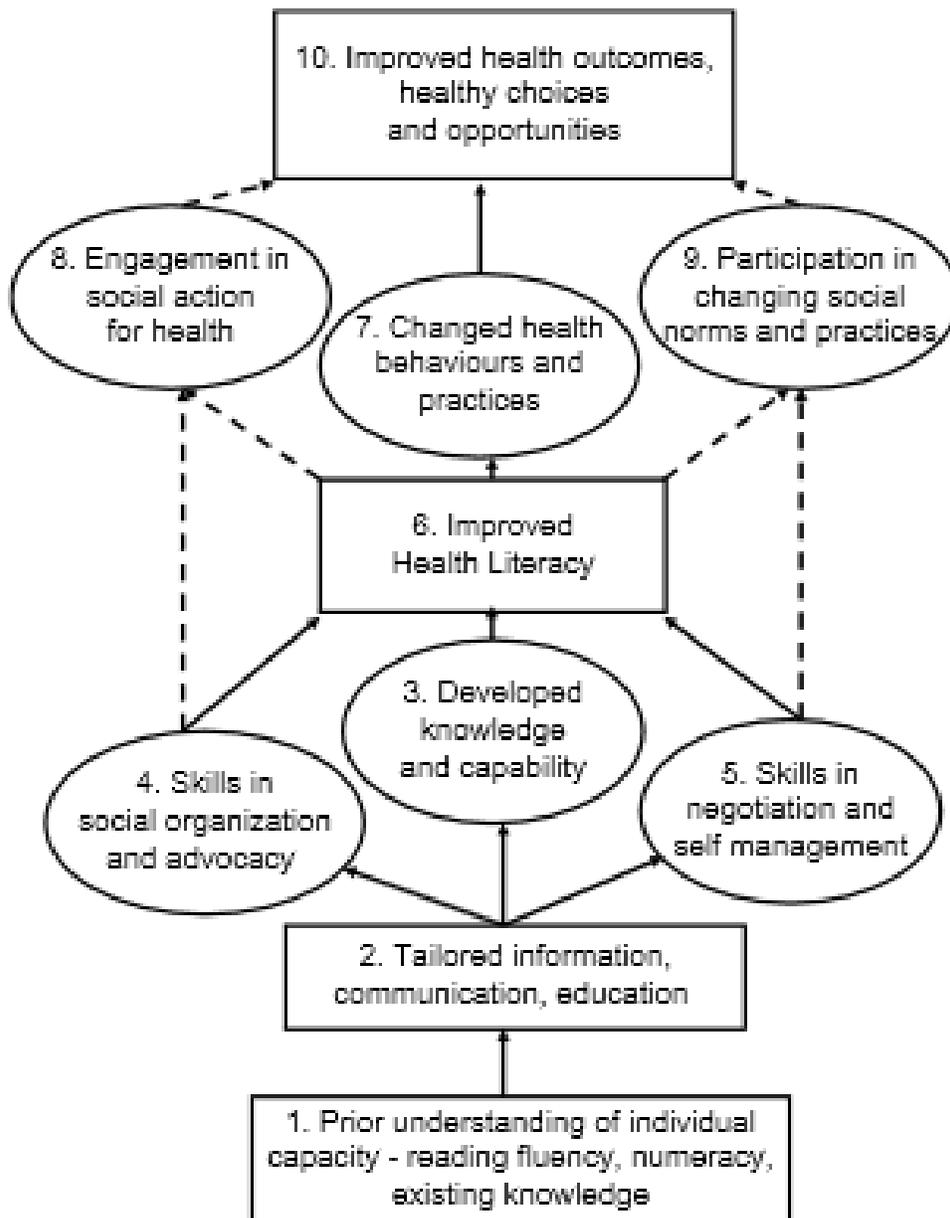
in health practices).

Figure 5: Diagram of Asset Model of Health Literacy From Nutbeam (2008).

Assuming we replace references to "health" with "education", we find this framing compelling as it shifts the discussion of the role of assessment, and appropriate aspects of assessment communication, to the larger goal structure of education. Although assessment-based communication should not "take on" the full range of enablement suggested here, it does raise questions regarding the expected role of assessment in the educational process, and what way we should consider assessment information in a larger ecosystem of educational information and educational literacy.

For example, while current norms for reporting assessment results highlight the need to make recommendations regarding actions that parents can undertake (e.g. Goodman & Hambleton, 2004) the literature is less prescriptive on how it is expected such information will influence the educational system or how information from assessment activity may interact with information regarding instructional standards or practices both inside and outside the educational setting. In this case, "educational literacy" refers to the large scale understanding of how learning outcomes are shaped and changed by practices.

While broadening the conversation to this macro level may seem a far departure from the topic of "score reporting", we believe shifts in technology related to data collection, analysis, and management (Behrens, Mislevy, DiCerbo, & Levy, 2012) are moving assessment systems toward integrated spaces with instructional system, thereby questioning both the boundaries of the disciplines and activities as well as the theoretic lenses appropriate to discourse in these areas. For example, recent multi-state consortium-based assessment systems were required to have strong alignment with instructional support and student information systems.

These shifts will likely require a conceptualization of assessment communication and assessment use in the larger educational governance and practice context likewise portrayed in the health literacy literature. In addition to the expansion and integration of

assessment concerns with instructional and institutional data, the health literacy/education literacy model raises questions regarding what knowledge and information is required for appropriate consumer and partner roles in the consumption and partnership of educational systems. What do parents and teachers need to know to properly consume assessment information as well as other types of educational information?   What are the boundaries between instructional and assessment information in the new technologically integrated world, and how will the assessment community respond to it?

Conclusion

This paper lays out three lenses through which to view the challenges of reporting assessment information. First, we discussed inputs to the assessment argument and communication given agreement on goals and motivations and conceptual frames. Second, we discussed a framework to consider variation in goals and interpretations and to consider the social context of system interaction. Third, we considered the hypothesized theories of action by which availability of student data would effect change in the education system. While it may seem daunting to consider these wide-ranging concerns in the design of a "simple" system to report results of a single test, we believe that the shifting political and technological landscape requires us to consider the demand for and utility of score reports as they are currently conceptualized. For some uses and stakeholders, reporting in this way may make sense. For other uses and stakeholders, a focus on skills may be more appropriate to support decisions that need to be made, and require integration of information across multiple sources. Still others may require things we won't know until we conduct analyses suggested by Activity Theory and educational literacy. We encourage those developing systems of reporting results of assessments of all kinds to understand and explore through the lenses outlined here.

References

Behrens, J. T., & DiCerbo, K. E. (2013). Technological Implications for Assessment Ecosystems: Opportunities for Digital Technology to Advance Assessment. Princeton, NJ: The Gordon Commission.

Behrens, J. T., Mislevy, R. J., DiCerbo, K. E., & Levy, R. (2012). Evidence centered design for learning and assessment in the digital world. In M. Mayrath, J. Clarke-Midura, D. H. Robinson, & G. Schraw (Eds.). T*echnology-based assessments for 21st Century skills: Theoretical and practical implications from modern research* (pp. 13-54). Charlotte, NC: Information Age Publishing.

Bradshaw, J., & Wheater, R. (2009, December). *International survey of results reporting*. Coventry, UK: Office of Qualifications and Examinations Regulation.

Engeström, Y., Miettinen, R., & Punamäki, R.-L. (1999). Perspectives on activity theory. Cambridge University Press.

Frezzo, D. C., Behrens, J. T., & Mislevy, R. J. (2009). Activity Theory and Assessment Theory in the Design and Understanding of the Packet Tracer Ecosystem. *International Journal of Learning and Media*, *1*(2). doi:10.1162/ijlm.2009.0015

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education,* 17(2), 145–220.

Hambleton, R. K. (2002). How can we make NAEP and state test score reporting scales and reports more understandable? In R. W. Lissitz & W. D. Schafer (Eds.), *Assessment in educational reform* (pp. 192-205). Boston: Allyn & Bacon.

Hambleton, R. K., & Slater, S. (1997). *Are NAEP executive summary reports understandable to policy makers and educators?* (CSE Technical Report 430). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Teaching.

Hattie, J. (2009, April). *Visibly learning from reports: The validity of score reports*. Paper

presented at the meeting of the National Council on Measurement in Education, San Diego, CA.

Jaeger, R. (1998). *Reporting the results of the National Assessment of Educational Progress* (NVS NAEP Validity Studies). Washington, DC: American Institutes for Research.

Jaeger, R. M. (2003). *NAEP validity studies: Reporting the results of the National Assessment of Educational Progress.* Working paper 2003–11. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Jonassen, D. H., Tessmer, M., & Hannum, W. H. (1999). *Task analysis methods for instructional design.* Psychology Press.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.

Lord, F. M. (1980). *Applications of Item Response Theory To Practical Testing Problems*. Routledge.

Mayrath, M., Clarke-Midura, J., Robinson, D. H. & Schraw, G. (Eds.). *Technology-based assessments for 21st Century skills: Theoretical and practical implications from modern research* (pp. 13-54). Charlotte, NC: Information Age Publishing.

Mislevy, R. J. (1994). Evidence and Inference in Educational Assessment. *Psychometrika*, *59*(4), 439–483.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3–67.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, *19*(4), 477–496. doi:10.1191/0265532202lt241oa

National Educational Goals Panel (NEGP). (1998). *Talking about tests: An idea book for state leaders*. Washington, DC: U.S. Government Printing Office.

Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine*, *67*(12), 2072–2078. doi:10.1016/j.socscimed.2008.09.050

Patelis, T. & Matos-Elefonte, H. (2009). *Efforts to produce relevant score reports to school, district, and state officials on national tests*. Paper presented at The National Council on Measurement in Education, San Diego, CA.

Ryan, J. M. (2006). Practices, issues, and trends in student test score reporting. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum.

Robinson, D. H., & Schraw, G. (1994). Computational efficiency through visual argument: Do graphic organizers communicate relations in text too effectively? *Contemporary Educational Psychology*, 19, 399-415.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, 91, 690-702.

Trout, D. L., & Hyde, E. (2006, April). *Developing score reports for statewide assessments that are valued and used: Feedback from K-12 stakeholders*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2010). An evidence-centered approach to using assessment data for policymakers. RR-10-03). Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/RR-10-03.pdf

Vezzu, M., VanWinkle, W., & Zapata-Rivera, D. (2012). *Designing and evaluating an interactive score report for students*. ETS Research Memorandum No. RM–12-01). ETS, Princeton.

Wainer,H., Hambleton,R.K.,& Meara,K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement*, *36*(4), 301–335.

Winn, W. (1988). Recall of the pattern, sequence, and names of concepts presented in instructional diagrams. *Journal of Research in Science Teaching*, 25, 375-386.

Zenisky, A. L., & Hambleton, R. K. (2012). Developing test score reports that work: The process

and best practices for effective communication. *Educational Measurement: Issues and

Practice*, *31*(2), 21-26.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2009). Getting the message out: An evaluation

of NAEP score reporting practices with implications for disseminating test results. *Applied

Measurement in Education*, *22*(4), 359–375.

Zapata-Rivera, D., & Zwick, R. (2011). *Test Score Reporting: Perspectives From the ETS Score

Reporting Conference*.

Latest edit: September 4, 2013