Tools and Tool-Makers for Future Educational Assessments?

Li Cai
CRESST/UCLA
Psychometrics?

Much of the technical work in educational assessment is psychometric in nature, or at least employs psychometric methods in some fashion.

Much of classical and contemporary research on psychometrics is based on principles from statistics, computer science, and engineering.

Not surprising if you think about the historical contexts in which other *-metrics disciplines developed within their own fields, e.g., econometrics, biometrics, …

And when our eyes glaze over mountains of assessment data, we look for simple answers from statistical analyses and hope (correctly or erroneously) that the assessment/tasks/items may be “blessed” by technical indicators of quality.
Psychometricians?

Thissen (2001) in his presidential address to the Psychometric Society noted that psychometricians study (mathematics, statistics, computing, etc.) “in order to build tools (largely) in order (for others) to study.” (p. 474).

I would argue that it remains (to a good approximation) a highly accurate characterization, given the ubiquitously tight assessment production timelines, the often carefully engineered process to make and maintain educational assessments, and the degree of commercialization.

I would also argue that this current state of affairs is less than desirable in a time of growing pressure on the why, how, and what of educational assessment.
Operational educational assessment, as a proper engineering problem, generates interesting applied research questions. On the other hand, fundamental research and development, occurring largely in academic and semi-academic settings have had a more difficult time penetrating into operation.

Certain things that would provide direct answers to some of the questions raised in the Gordon Commission report - things we understand that we now know how to do (efficiently enough for operational purposes) - have remained the next wave that never reaches the shore.

Meaningful feedback
Comparability
Adaptivity
accountability
Scientists that Occasionally Foray into Engineering

Most of the engineering in psychometrics involves designing tools (software) for data analysis and modeling.

Tools that can be used directly in operation; not a scientist’s own (mostly) kludgy one-time purposes.

This requires both broad conceptual frameworks, carefully designed reusable resources, and a platform for implementation that end-users can learn.

I mention Darrell Bock because he is an example of a true scientist (and artist) that ventured into tools-making, first for himself (and his students) to study, and then mostly for others.
L. L. Thurstone himself provided a great example with his (1927) work on law of comparative judgment.

More recently Bob Mislevy and colleagues’ work on ECD

Perhaps not to become subject matter experts, but at least sufficiently proficient with the major trends to understand what problems are worth spending time on.

The emphasis on assessment design highlights the reason why psychometricians cannot sit idle until pilot/field testing data become available.

Psychometricians ought to work more productively with other team members…and not just be that “numbers person.”
In encouraging students to help us build an integrated interpretation of mental phenomena on an experimental foundation, let us remember that a psychological theory is not good simply because it is cleverly mathematical, that an experiment is not good just because it involves ingenious apparatus, and that statistics are merely the means for checking theory with experiment.