Topic modeling is applied with science content standards to evaluate semantic clustering. The probability that each item from a statewide assessment belongs to each cluster/topic is then estimated as a source of content-related validity evidence. We also show how visualizations can map the content coverage of the test.

**Lit Review**

(Additions March 2019 by SS)

DUE MARCH 18 to Discussant!

In the last decade, the landscape of education has evolved through the passage of legislation aimed at improving student outcomes through standards driven accountability such as No Child Left Behind and current day Every Student Succeeds Act [@beckerbuckendahl2013]. This has led to a vast increase in standardized educational assessment use. As part of this movement an intensive focus on alignment has also emerged due to the need for congruence between assessment, standards, curricula, and instruction in order to impact student outcomes (e.g. Porter, 2002). Specifically, the goal “of alignment is to create a coherent educational system that conveys a clear and unified message about expectations and goals” [@vockleylang2009, p.8].

Content-related validity is a critical component of the "overall evaluative judgment" [@messick95, p. 741] of the validity of test scores for a given use and is one of the five major sources of validity evidence outlined by the \*Standards for Educational and Psychological Testing\* [@standards14]. Empirical evaluations of content related validity evidence generally come in the form of alignment studies, with panels of experts judging the alignment between the content represented in the standards and the content represented in the test items [@sireci07; @webb97]. This process can be time consuming and cost intensive [@anderson15], therefore, this paper aims to add to the literature on methodological advancements in content-related validity, not aiming to replace but rather supplement information from alignment studies through efficient but effective means. We propose using machine-learning based approaches, specifically topic modeling, to mine the educational standards for groups of words that co-occur frequently. These data-derived clusters can then be evaluated for substantive meaning, and the extent to which the text within individual test items corresponds with these clusters can be directly evaluated as an additional, alternative source of content-related validity evidence.

Topic modeling is a statistical method for garnering meaningful insight from data [@morhbogdanov2013]. Specifically, it is a probabilistic method for identifying latent topics in text based documents. It has a history stemming from content analysis and latent semantic analysis but instead of the researcher predetermining the topics to be analyzed and coded, the topics emerge from corpora of text. Topic modeling has advanced the field of text analysis from simply identifying specified words through a deductive approach where topics are pre-identified to a more inductive approach where meaning is allowed to emerge through a corpus of texts [@morhbogdanov2013]. Topic modeling is still relatively new in the scheme of text-based analytic research with the first article by Blei published in 2003 [@beli2003]. Prior to these advancements in machine learning and more specifically topic modeling, inductive themes or latent meanings in text were achievable mainly through more traditional qualitative analysis [@nikolenko2017]. Given recent advancements in machine learning, topic modeling is demonstrating potential for use in a broad array of contexts with good success including use in business for sentiment analysis of consumer comments about products and services [@lin2009], understanding political themes across a range of documents (e.g. Hagen, 2018), and even aggregating results across scientific studies despite differences in terminology and fields of inquiry [@geffen2017]. Topic modeling has even shown emerging potential to produce similar results to traditional qualitative frameworks such as grounded theory on qualitative survey data when utilizing a semi-supervised form of LDA [@nikolenko2017]. The increasing accuracy of such techniques thereby provide support for potential use in expanded applications, such as that proposed in this paper.

Overall, in this paper we explore the use of topic modeling and specifically latent dirichlet analysis (LDA) as part of the alignment process in content-related validity. We adopt a similar theoretical framework to traditional approaches to content-related validity, but apply topic modeling procedures to evaluate the correspondence between the language used in content standards and the language used in the test items (i.e., the item stems and response option as an additional source of content-related validity evidence). Although there are limitations in the application of topic modeling and specifically LDA requiring content knowledge of the researchers and understanding of machine learning to determine appropriate technique and level of supervision of the model, we propose that it does hold potential as a triangulated source of validation in the process of establishing construct related validity.