

1116-L1-1124

James H. Fife* (jfife@ets.org), Educational Testing Service, 660 Rosedale Road, Princeton, NJ 08541. *Modeling the Difficulty of Constructed-Response Items.*

The advantages of automatic item generation over the traditional hand-crafting of individual items have been well-documented (Bejar et al., 2003). The automatic generation of items can reduce test development costs and, when item models are written as part of an evidence centered design approach, can increase validity (Mislevy, Steinberg, & Almond, 2003). Additionally, when the factors that determine difficulty are carefully controlled, items can be generated at prescribed difficulty levels (Graf & Fife, 2013), perhaps making possible the calibration of the item model rather than the calibration of the individual items, further reducing costs. In this talk, I will describe a project in which ten item models were created, using disclosed GRE items as parent items, with an attempt to control for difficulty. Five items were generated from each model; these child items were piloted along with the parent items on a self-selected population of GRE registrants. The results were used to determine the difficulty of the child items, compared with each other and compared with the parent item. For nine of the models, the variants had similar difficulty and discrimination, but for one of the models, there were non-construct-related factors that resulted in items of varying difficulty. (Received September 17, 2015)