Appendix 1. Item Response Theory Graded Response Model.

This GRM model was fit to the data to estimate location and information parameters for each item.

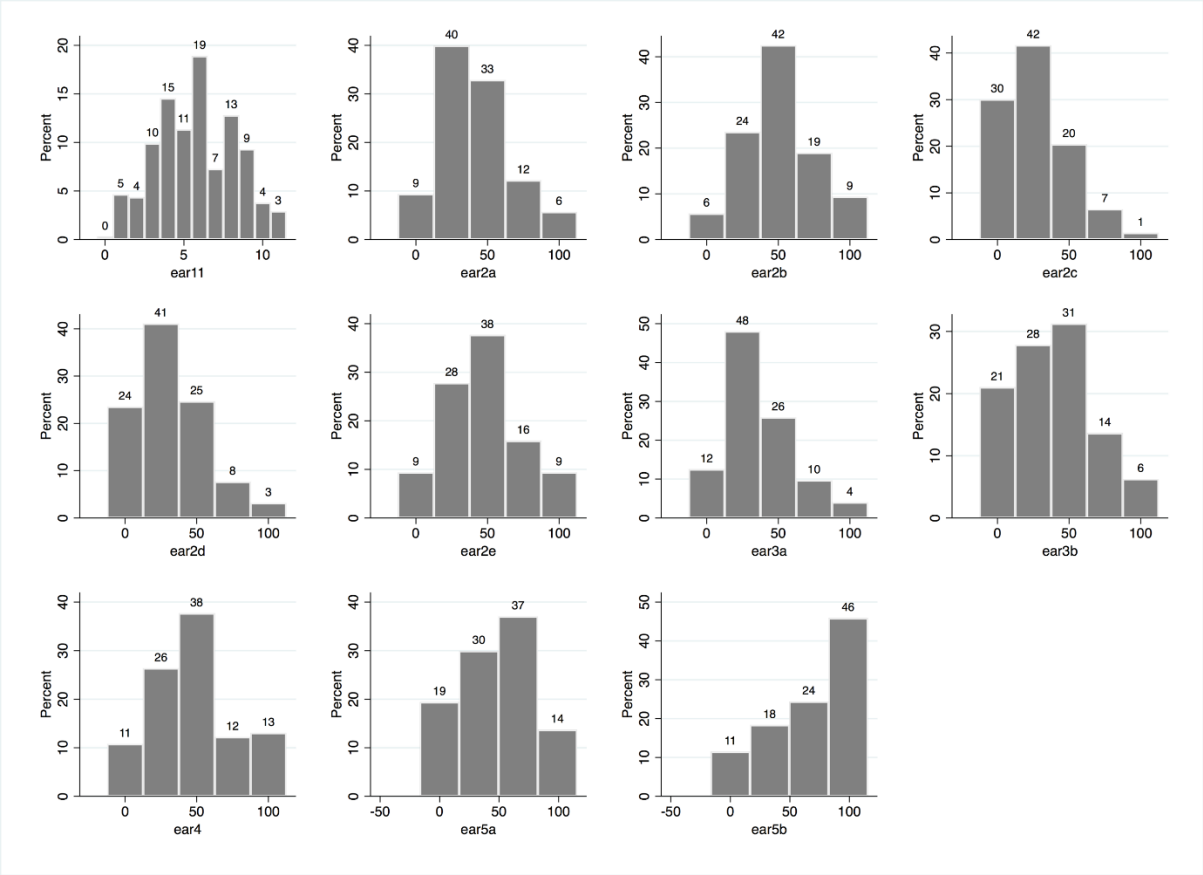
The GRM is specified as follows:

$$P\left(X_{pi} \geq k \middle| a_i, b_{ik}, \theta_p\right) = \frac{1}{\left(1 + \exp\left(-a_i\left(\theta_p - b_{ik}\right)\right)\right)};\ \theta_p \sim N(0,1)$$

Here, *X* represents patient p's response to each item *i.* The GRM models the probability of a patient selecting a score at or above each item score category. With a five-category Likert-scaled item scored 1-5, for example, the probability of scoring at or above 1 is trivially 1, but for score points $k = 2 \ldots 5$, the probability is modeled as four logistic functions rising from 0 to 1 across the latent score scale, θ. Here, θ_p represents the latent score for each patient p (hypothesized to indicate a patient's perceived ability to hear), a_i represents the information parameter for each item i, and b_{ik} indicates the location parameter for each item i and score category k. The information parameter (a_i) indicates how well an item can distinguish between patients with very similar latent abilities. The location parameter (b_{ik}) indicates whether patients need a higher or lower level of perceived hearing ability, θ_p, to respond at or above that level k. The θ_p parameters can be interpreted on a standard normal scale, where -1 and $+1$ are one standard deviation below and above the mean, respectively. Direct interpretation of IRT parameter values can be cumbersome. Strictly speaking, a_i is the increase in the log of the odds of scoring at or above item i's categories k, for each 1-standard-deviation-unit increase in θ_p. The location parameters, b_{ik}, indicate where a patient with $\theta_p = b_{ik}$ has a 50% chance of scoring at or above category k of item i. These parameters are theoretically invariant to populations of

items and examinees. The figure below displays item-level histograms. These help to evaluate whether response categories for each item have a sufficient number of patient responses to estimate corresponding location parameters b_{ik}.

Appendix 1 Figure: Item-score histograms for items ear11-ear5b (n=353).



Appendix 2. Category characteristic curves for all 11 items, estimated from a Graded Response Model (n=353)