

Workshop 2: Modeling the effects of collaboration on student performance

Peter F. Halpin

April 19, 2015



NEW YORK UNIVERSITY

Outline

Part 1: How to define and measure the effects of collaboration on academic performance?

- ▶ Review of research on small groups
- ▶ A framework based on IRT
- ▶ Review of some results from pilot data

Outline

Part 1: How to define and measure the effects of collaboration on academic performance?

- ▶ Review of research on small groups
- ▶ A framework based on IRT
- ▶ Review of some results from pilot data

Part 2: Implications for assessment design

- ▶ Some results on discrimination and ability in the 2PL model
- ▶ Initial considerations about IRFs for “new” item types

Outline

Part 1: How to define and measure the effects of collaboration on academic performance?

- ▶ Review of research on small groups
- ▶ A framework based on IRT
- ▶ Review of some results from pilot data

Part 2: Implications for assessment design

- ▶ Some results on discrimination and ability in the 2PL model
- ▶ Initial considerations about IRFs for “new” item types

Part 3: Exercises in R; code and data available at

- ▶ `devtools::install_github('peterhalpin/BearShare')`
- ▶ Also see git repo for example code: [Collab_Outcomes.html](#)

Part 1: A basic scenario for collaboration

- ▶ Two students each write a conventional math assessment
- ▶ Their math ability is estimated to be θ_j and θ_k
- ▶ The two students then work together on a second conventional math assessment
- ▶ What do we expect about their performance on the second test, based on the first?

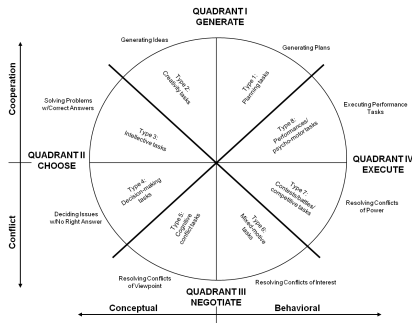
Collaboration as a social psychological question

- ▶ Research on group problem solving dates back to Shaw (1932)
- ▶ Selective review of some precedents for a psychometric theory of collaboration

Task types

Intellective tasks:

- ▶ Defined as having a demonstrably “correct” answer with respect to an agreed upon system of knowledge
- ▶ Differentiated from decision / judgement tasks on a continuum of *demonstrability* (Laughlin 2011)
- ▶ Differentiated from mixed-motive tasks in that the goals and outcomes are the same for all members.



McGrath's (1984) group task circumplex

Lorge & Solomon 1955¹

Model A

Under Model A the probability of a group solution is the probability that the group contains one or more members who can solve the problem. This non-interactional ability model for any specific problem can be expressed mathematically as follows: Let

P_g = the probability that a group of size k solve the problem;

P_I = the probability that an individual solve the problem.

Then

$$P_g = 1 - (1 - P_I)^k, \quad (2)$$

where P_g and P_I are population parameters considered fixed for the specific problem and the specific population.

¹Two models of group behavior in the solution of Eureka-type problems. *Psychometrika*, 1955, 20 (2), p. 141

Lorge & Solomon 1955²

Model A

Under Model A the probability of a group solution is the probability that the group contains one or more members who can solve the problem. This non-interactive ability model for any specific problem can be expressed mathematically as follows: Let

P_g = the probability that a group of size k solve the problem;

P_I = the probability that an individual solve the problem.

Then

$$P_g = 1 - (1 - P_I)^k, \quad (2)$$

where P_g and P_I are population parameters considered fixed for the specific problem and the specific population.

²Two models of group behavior in the solution of Eureka-type problems. *Psychometrika*, 1955, 20 (2), p. 141

Smoke and Zajonc 1962³

If p is the probability that a given individual member is correct, the group has a probability $h(p)$ of being correct, where $h(p)$ is a function of p depending upon the type of decision scheme accepted by the group. We shall call $h(p)$ a decision function. Intuitively, it would seem that a decision scheme is desirable to the extent that it surpasses p .

³On the reliability of group judgements and decisions. In *Mathematical methods for small group processes* (Eds. Criswell, Solomon, Suppes), p. 322

Resources constitute “all the relevant knowledge, abilities, skills, or tools possessed by the individual(s) who is attempting to perform a task” (Steiner, 1966, p. 274).

Transformers constitute all the variables that have an impact on resources and determine the manner in which they are incorporated into and related to the output variables. Transformers include such variables as situational and task constraints, role systems, and certain personal characteristics that may affect the way personal task-relevant resources are utilized in the output.

To summarize, input and output variables have been categorized in a manner that allows the model to be stated, in its simplest form, as $P = f(T, R)$, where P represents the group output or product, T stands for transformer variables, and R represents resource variables.

⁴Towards a general model of group productivity. *Psychological Bulletin*, 86 (1), pp. 67-68

Summary

- ▶ Building on research on small groups:
 - ▶ Intellective tasks (vs decision tasks)
 - ▶ Cooperative group interactions (vs competitive or mixed-motive)
 - ▶ Describing group outcomes via decision / functions that depend on characteristics of individuals
- ▶ But with a focus on:
 - ▶ Letting probability of success vary over individuals (e.g., via ability)
 - ▶ Describing relevant task characteristics (e.g., via difficulty)
 - ▶ The performance of individual groups rather than groups in aggregate

Back to our basic scenario

- ▶ Two students each write a conventional math assessment
- ▶ Their math ability is estimated to be θ_j and θ_k
- ▶ The two students then work together on a second conventional math assessment
- ▶ What do we expect about their performance on the second test, based on the first?

Collaboration as a psychometric question

- ▶ Traditional psychometric models assume conditional independence of the items

$$p(\mathbf{x}_j \mid \theta_j) = \prod_i^N p(x_{ij} \mid \theta_j) \quad (1)$$

- ▶ Traditional psychometric models also assume that the responses of two (or more) persons are independent

$$p(\mathbf{x}_j \mathbf{x}_k \mid \theta_j \theta_k) = p(\mathbf{x}_j \mid \theta_j) p(\mathbf{x}_k \mid \theta_k) \quad (2)$$

- ▶ When people work together does equation (2) hold?

“Working together” in terms of scoring rules⁵

- ▶ For binary items and pairs of responses, consider:

- ▶ The conjunctive rule

$$x_{ijk} = \begin{cases} 1 & \text{if } x_{ij} = 1 \text{ and } x_{ik} = 1 \\ 0 & \text{otherwise} \end{cases}$$

- ▶ The disjunctive rule

$$x_{ijk} = \begin{cases} 0 & \text{if } x_{ij} = 0 \text{ and } x_{ik} = 0 \\ 1 & \text{otherwise} \end{cases}$$

- ▶ More possibilities, especially for items with > 2 responses or groups with > 2 collaborators

⁵cf. Steiner's 1966 classification of task types

Scoring rules vs decision functions

- ▶ Scoring rules describe what “counts” as a correct group response
 - ▶ Under control of the test designer⁶
- ▶ Decision functions describe the strategies adopted by a team
 - ▶ Under control of the team

⁶ Maris & van der Maas (2012). Speed-accuracy response models: scoring rules based on response time and accuracy. *Psychometrika*, 77 (4), 615-633

Scoring rules vs decision functions

- ▶ Scoring rules describe what “counts” as a correct group response
 - ▶ Under control of the test designer⁶
- ▶ Decision functions describe the strategies adopted by a team
 - ▶ Under control of the team
- ▶ Basic research strategy
 - ▶ Assume a certain scoring rule
 - ▶ Consider plausible models for team strategies
 - ▶ Test the models against data

⁶Maris & van der Maas (2012). Speed-accuracy response models: scoring rules based on response time and accuracy. *Psychometrika*, 77 (4), 615-633

“Working together” in terms of scoring rules

- ▶ For binary items and pairs of responses, consider:

- ▶ The conjunctive rule

$$x_{ijk} = \begin{cases} 1 & \text{if } x_{ij} = 1 \text{ and } x_{ik} = 1 \\ 0 & \text{otherwise} \end{cases}$$

- ▶ The disjunctive rule

$$x_{ijk} = \begin{cases} 0 & \text{if } x_{ij} = 0 \text{ and } x_{ik} = 0 \\ 1 & \text{otherwise} \end{cases}$$

- ▶ More possibilities, especially for items with > 2 responses or groups with > 2 collaborators

Defining successful pairwise collaboration

- The independence model

$$E_{\text{ind}}[x_{ijk} \mid \theta_j \theta_k] = E[x_{ij} \mid \theta_j] E[x_{ik} \mid \theta_k]$$

Defining successful pairwise collaboration

- ▶ The independence model

$$E_{\text{ind}}[x_{ijk} \mid \theta_j \theta_k] = E[x_{ij} \mid \theta_j] E[x_{ik} \mid \theta_k]$$

- ▶ Successful collaboration

$$E[x_{ijk} \mid \theta_j \theta_k] > E_{\text{ind}}[x_{ijk} \mid \theta_j \theta_k]$$

- ▶ Unsuccessful collaboration

$$E[x_{ijk} \mid \theta_j \theta_k] < E_{\text{ind}}[x_{ijk} \mid \theta_j \theta_k]$$

- ▶ Note: these definitions are item- and dyad- specific

Some models for successful collaboration

- ▶ Minimum individual performance (disruptive team member)

$$E_{\min}[x_{ijk} \mid \theta_j \theta_k] = \min\{E[x_{ij} \mid \theta_j], E[x_{ik} \mid \theta_k]\}$$

Some models for successful collaboration

- ▶ Minimum individual performance (disruptive team member)

$$E_{\min}[x_{ijk} \mid \theta_j \theta_k] = \min\{E[x_{ij} \mid \theta_j], E[x_{ik} \mid \theta_k]\}$$

- ▶ Maximum individual performance (cheating / tutor)

$$E_{\max}[x_{ijk} \mid \theta_j \theta_k] = \max\{E[x_{ij} \mid \theta_j], E[x_{ik} \mid \theta_k]\}$$

Some models for successful collaboration

- ▶ Minimum individual performance (disruptive team member)

$$E_{\min}[x_{ijk} \mid \theta_j \theta_k] = \min\{E[x_{ij} \mid \theta_j], E[x_{ik} \mid \theta_k]\}$$

- ▶ Maximum individual performance (cheating / tutor)

$$E_{\max}[x_{ijk} \mid \theta_j \theta_k] = \max\{E[x_{ij} \mid \theta_j], E[x_{ik} \mid \theta_k]\}$$

- ▶ “True collaboration”

$$E[x_{ijk} \mid \theta_j \theta_k] \geq \max\{E[x_{ij} \mid \theta_j], E[x_{ik} \mid \theta_k]\}$$

A model for “true collaboration”

- An additive model

$$E_{\text{add}}[x_{ijk} \mid \theta_j \theta_k] = E[x_{ij} \mid \theta_j] + E[x_{ik} \mid \theta_k] - E[x_{ijk} \mid \theta_j \theta_k]$$

A model for “true collaboration”

- ▶ An additive model

$$E_{\text{add}}[x_{ijk} \mid \theta_j \theta_k] = E[x_{ij} \mid \theta_j] + E[x_{ik} \mid \theta_k] - E[x_{ijk} \mid \theta_j \theta_k]$$

- ▶ Recalling $E[x_{ijk} \mid \theta_j \theta_k] > E[x_{ij} \mid \theta_j]E[x_{ik} \mid \theta_k]$, define an additive independence (AI) model

$$\begin{aligned} E_{AI}[x_{ijk} \mid \theta_j \theta_k] &= E[x_{ij} \mid \theta_j] + E[x_{ik} \mid \theta_k] - E[x_{ij} \mid \theta_j] E[x_{ik} \mid \theta_k] \\ &\geq E_{\text{add}}[x_{ijk} \mid \theta_j \theta_k] \end{aligned}$$

- ▶ AI is an upper bound on any “more interesting” additive model for successful collaboration

More on AI model

- Can also be written as:

$$\begin{aligned} E_{AI}[x_{ijk} \mid \theta_j \theta_k] &= E[x_{ij} \mid \theta_j] (1 - E[x_{ik} \mid \theta_k]) \\ &\quad + E[x_{ik} \mid \theta_k] (1 - E[x_{ij} \mid \theta_j]) \\ &\quad + E[x_{ij} \mid \theta_j] E[x_{ik} \mid \theta_k] \end{aligned}$$

- Which has an interpretation in terms of three cases

More on AI model

- And is also equivalent to Lorge & Solomon's Model A

$$E_{AI}[x_{ijk} \mid \theta_j \theta_k] = 1 - (1 - E[x_{ij} \mid \theta_j])(1 - E[x_{ik} \mid \theta_k])$$

- Except the “probability an individual can solve the problem” now depends on both the individual and the problem

More on AI model⁷

- We probably want some constraints on what counts as a good collaborative IRF

2.2. *Latent monotonicity.* We say that a latent variable model satisfies the condition of *latent monotonicity* if the functions

$$1 - F_j(x|\mathbf{u}) = P(X_j > x | \mathbf{U} = \mathbf{u})$$

are all nondecreasing functions of \mathbf{u} for all values of x and for $j = 1, \dots, J$. In case \mathbf{U} is a vector, latent monotonicity requires that $1 - F_j(x|\mathbf{u})$ be nondecreasing in each coordinate of \mathbf{u} .

⁷Holland & Rosenbaum (1986). Conditional Association and Unidimensionality in Monotone Latent Variable Models. *The Annals of Statistics*, 14 (4), 1523 – 1543

More on AI model⁷

- We probably want some constraints on what counts as a good collaborative IRF

2.2. *Latent monotonicity.* We say that a latent variable model satisfies the condition of *latent monotonicity* if the functions

$$1 - F_j(x|\mathbf{u}) = P(X_j > x | \mathbf{U} = \mathbf{u})$$

are all nondecreasing functions of \mathbf{u} for all values of x and for $j = 1, \dots, J$. In case \mathbf{U} is a vector, latent monotonicity requires that $1 - F_j(x|\mathbf{u})$ be nondecreasing in each coordinate of \mathbf{u} .

- Easy to show that AI satisfies latent monotonicity, if the individual IRFs do (trivial for other models also)

⁷Holland & Rosenbaum (1986). Conditional Association and Unidimensionality in Monotone Latent Variable Models. *The Annals of Statistics*, 14 (4), 1523 – 1543

AI: latent monotonicity

Assumptions:

$$f(x) \geq f(x') \text{ for } x > x' \quad \text{and} \quad 0 \leq g(y) \leq 1$$

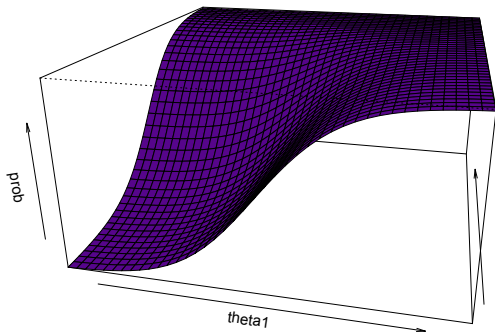
Show:

$$f(x) + g(y) - f(x) g(y) \geq f(x') + g(y) - f(x') g(y)$$

Contradiction:

$$\begin{aligned} f(x) + g(y) - f(x) g(y) &< f(x') + g(y) - f(x') g(y) \\ \rightarrow f(x) - f(x') &< g(y) (f(x) - f(x')) \end{aligned}$$

AI: example IRF⁸



⁸Using 2PL model for individual IRFs with $\alpha = 1$ and $\beta = 0$

Models abound!

- ▶ Basic idea: write down IRFs for collaboration based on assumed-to-be-known individual abilities (and item parameters)
- ▶ But how do we characterize empirical team performance?

Empirical team performance

- ▶ We have
 - ▶ Observed collaborative responses $\mathbf{x}_{jk} = (x_{1jk}, x_{1jk}, \dots, x_{mjk})$
 - ▶ A model for individual performance on the m (conventional) math items

Empirical team performance

- ▶ We have
 - ▶ Observed collaborative responses $\mathbf{x}_{jk} = (x_{1jk}, x_{1jk}, \dots, x_{mjk})$
 - ▶ A model for individual performance on the m (conventional) math items
- ▶ So we can get “team theta,” e.g.,

$$\hat{\theta}_{jk} = \operatorname{argmax}_{\theta} \{L_0(\mathbf{x}_{jk} \mid \theta)\} \quad (3)$$

- ▶ Where L_0 is the likelihood of the model calibrated on individual performance (reference model)

Proposed method for testing models

- ▶ Testing of different models against reference model

$$D_{\text{model}} = -2 \ln \frac{L_{\text{model}}(\mathbf{x}_{jk} \mid \theta_j \theta_k)}{L_0(\mathbf{x}_{jk} \mid \hat{\theta}_{jk})} \quad (4)$$

- ▶ Also a “direct test” of effect of collaboration for each individual

$$D_0 = -2 \ln \frac{L_0(\mathbf{x}_{jk} \mid \theta_j)}{L_0(\mathbf{x}_{jk} \mid \hat{\theta}_{jk})} \quad (5)$$

with effect size $\delta_{jk} = \frac{\theta_{jk} - \theta_j}{\sigma_\theta}$

Proposed method: caveat 1

- ▶ Reference distribution for D_{model}
 - ▶ Ind and AI models are not nested with reference model \rightarrow No Wilk's theorem
 - ▶ Can use Vuong's 1989⁹ results for LR with non-nested models, but asymptotic in m
 - ▶ Good news: we can bootstrap a null distribution for (4) and (5) pretty easily

⁹ Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrika*, 57(2), 307 – 333.

Bootstrapping the reference distribution

Assuming known item parameters and θ_j, θ_k . For $r = 1, \dots, R$

Step 1 Generate collaborative response patterns $\mathbf{x}_{jk}^{(r)}$ from $E_{\text{model}}[x_{ijk} \mid \theta_j \theta_k]$

Step 2 Compute $L_{\text{model}}(\mathbf{x}_{jk}^{(r)} \mid \theta_j \theta_k)$

Step 2 Estimate $\theta_{jk}^{(r)}$ for each $\mathbf{x}_{jk}^{(r)}$; save $L_0(\mathbf{x}_{jk}^{(r)} \mid \theta_{jk})$

Step 4 Compute $D_{\text{model}}^{(r)}$ or $D_0^{(r)}$

Proposed method: caveat 2

- ▶ Sampling error

- ▶ In practice we don't know the item parameters (estimation error) or θ_j , θ_k (prediction error)
- ▶ Could build these sources of error into bootstrap à la plausible values

Step 0 For each r , sample items from asymptotic distribution of items; θ_j , θ_k from their asymptotic distribution (with parameters obtained via resampling on *those* items)

Proposed method: caveat 2

- ▶ Sampling error

- ▶ In practice we don't know the item parameters (estimation error) or θ_j , θ_k (prediction error)
- ▶ Could build these sources of error into bootstrap à la plausible values

Step 0 For each r , sample items from asymptotic distribution of items; θ_j , θ_k from their asymptotic distribution (with parameters obtained via resampling on *those* items)

- ▶ It's not all that bad – item parameters pre-calibrated, θ_j and θ_k are predicted on a different set of items
 - ▶ The only thing estimated from \mathbf{x}_{jk} is θ_{jk}

Proposed method: caveat 3

- ▶ Estimating θ_{jk}
 - ▶ What guarantee do we have that actual, observed collaborative response patterns aren't "abberant" with respect to individually calibrated items?
 - ▶ None – but we don't have that for individual response patterns either!
 - ▶ Can assess person fit in usual ways for \mathbf{x}_{jk} as well

Proposed method: caveat 3

- ▶ Estimating θ_{jk}
 - ▶ What guarantee do we have that actual, observed collaborative response patterns aren't "abberant" with respect to individually calibrated items?
 - ▶ None – but we don't have that for individual response patterns either!
 - ▶ Can assess person fit in usual ways for \mathbf{x}_{jk} as well
- ▶ θ_{jk} via ML? For shame! MAP, EAP, WMLE, PV ...
 - ▶ But what is the prior in the collaborative condition? Need a latent regression / multi-group model to estimate $E(\theta_{jk})$

Example 1: friends dataset in BearShare.R

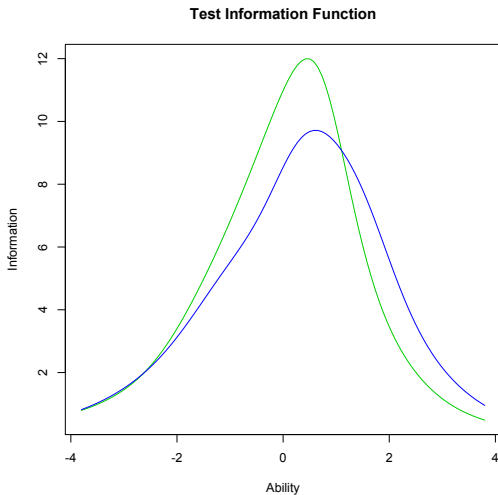
- ▶ Minimal design:
 - ▶ Pool of pre-calibrated math items (grade 12 NAEP, modified to be numeric response)
 - ▶ Individual “pre-test” → estimate individual abilities
 - ▶ Collaborative “post-test” → evaluate models, estimate δ_{jk}
 - ▶ Modality of collaboration: face-to-face
- ▶ Limitations:
 - ▶ Small calibration sample; crowd workers
 - ▶ Individual and collaborative forms were not counterbalanced (neither in order nor content)

A horizontal line with four points labeled M, N, P, and Q from left to right. Above the line, the distances between consecutive points are marked: 20 cm between M and N, 12 cm between N and P, and 16 cm between P and Q.

AMT Crowdworkers Demographic Survey

Variable	Levels	n	%	\sum %
Gender	Female	155	46.5	46.5
	Male	178	53.5	100.0
Age	18-30	117	35.2	35.2
	30-40	129	38.9	74.1
	40-55	71	21.4	95.5
	55+	15	4.5	100.0
Education	Some Grade School	3	0.9	0.9
	High School Diploma	49	14.7	15.6
	Some College	118	35.4	51.0
	Bachelor's Degree	132	39.6	90.7
	Master's Degree	22	6.6	97.3
	Ph.D or Advanced Degree	9	2.7	100.0
Country	United States	313	94.0	94.0
	India	16	4.8	98.8
	Canada	3	0.9	99.7
	United Kingdom	1	0.3	100.0
English First Lang	Yes	321	96.4	96.4
	No	12	3.6	100.0

Individual and collaborative test forms



Seven volunteers and their + 1

Instructions

On the [Courseware](#) page you will find two tests, Individual and Collaborative.

Please do the Individual test first, working alone (this is important!). Then do the Collaborative test with a partner. You can collaborate in person, over Skype, on the phone, or however you like.

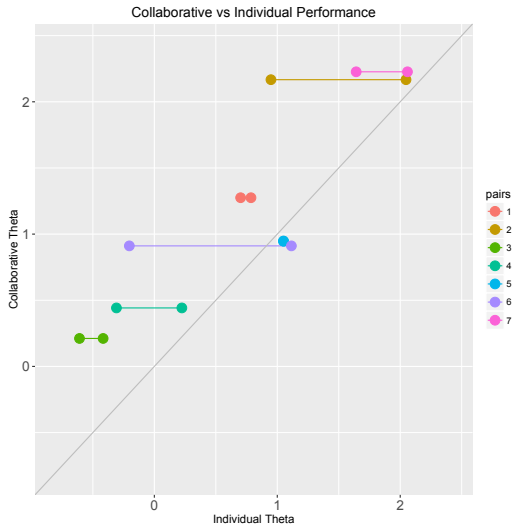
The test questions require numerical responses. Fractions can be written using $\frac{1}{2}$, but no letters are permitted. Once you have decided on an answer, press the "check" button and navigate to the next question.

To answer the questions, you can use the calculator widget that appears at the bottom of the test, or the internet. You may also want to use paper and a pencil.

Don't spend too long on any one question, but please try to provide an answer to all of the questions. If you have to guess, that is OK. Also, if you and your partner do not come to the same conclusion about a question, it is OK to give different answers.

Your participation in this project helps us to improve online math education. Thanks!

Deltas



Model tests: Sanity check using individual pre-test

Pairs	Ind	p_obs	Min	p_obs	Max	p_obs	AI	p_obs
1	-1.43	0.86	1.96	0.15	2.76	0.08	13.79	0.00
2	0.23	0.55	0.00	1.00	7.63	0.00	14.92	0.00
3	-2.26	0.85	0.98	0.32	2.28	0.10	8.70	0.00
4	-0.35	0.56	0.62	0.39	4.73	0.03	9.39	0.00
5	-0.23	0.59	5.89	0.01	5.90	0.02	28.74	0.00
6	1.03	0.31	0.06	0.80	18.37	0.00	26.84	0.00
7	0.69	0.38	0.12	0.71	1.59	0.22	10.69	0.00

Table reports D_{model} and $P(D > \text{obs})$ for individual pre-test scored using conjunctive scoring rule

Model tests: Collaborative data

Pairs	Ind	p_obs	Min	p_obs	Max	p_obs	AI	p_obs
1	15.63	0.00	2.88	0.09	2.08	0.15	0.90	0.33
2	13.03	0.00	9.15	0.01	0.07	0.82	2.09	0.16
3	21.30	0.00	5.87	0.00	3.50	0.06	0.33	0.41
4	15.36	0.00	5.22	0.02	0.46	0.53	0.85	0.30
5	6.91	0.01	0.10	0.74	0.10	0.74	8.09	0.00
6	19.21	0.00	11.78	0.00	0.38	0.52	3.30	0.06
7	6.45	0.00	1.71	0.23	0.12	0.69	5.29	0.00

Table reports observed D_{model} and $P(D > \text{obs})$ for collaborative test scored using conjunctive scoring rule

Model tests: Collaborative data

Pairs	Ind	p_obs	Min	p_obs	Max	p_obs	AI	p_obs
1	15.63	0.00	2.88	0.09	2.08	0.15	0.90	0.33
2	13.03	0.00	9.15	0.01	0.07	0.82	2.09	0.16
3	21.36	0.00	5.22	0.02	0.46	0.53	0.85	0.30
4	15.36	0.00	5.22	0.02	0.46	0.53	0.85	0.30
5	6.91	0.01	0.10	0.74	0.10	0.74	8.09	0.00
6	19.21	0.00	11.78	0.00	0.38	0.52	3.30	0.06
7	6.45	0.00	1.71	0.23	0.12	0.69	5.29	0.00

How distinct are these models ?

Table reports observed D_{model} and $P(D > \text{obs})$ for collaborative test scored using conjunctive scoring rule

Part 1: denouement

- ▶ Can define, estimate, and test models of collaboration on academic performance using IRT-based methods
- ▶ But how distinct are these models, really?

Part 2: assessment design

- ▶ In some cases the four original models are trivially equivalent
 - ▶ e.g., Min and Max models are equivalent when $\theta_j = \theta_k$
- ▶ Implications for designing collaborative assessments
 - ▶ Selecting items & matching partners to avoid cases where models are equivalent
- ▶ This part: stating model equivalencies in terms of 2PL

Results about item discrimination: Ind and Min models

Letting $\xi = \min\{\theta_j, \theta_k\}$, a lower bound on the IRF of the ind model:

$$P_i(\theta_j) P_i(\theta_k) \geq P_i(\xi)^2. \quad (6)$$

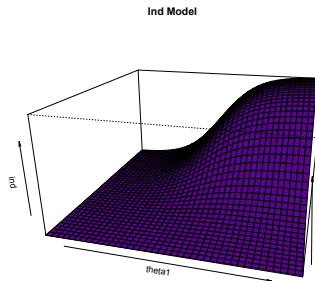
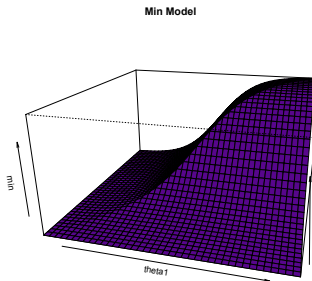
Then

$$\Delta_i(\xi) \equiv P_i(\xi) - P_i(\xi)^2 = \frac{1}{\alpha_i} \frac{\partial}{\partial \xi} P_i(\xi), \quad (7)$$

is an upper bound on the difference between min IRF and the independence IRF, leading to

$$\int_{-\infty}^{\infty} \Delta_i(\xi) d\xi = \frac{1}{\alpha_i}. \quad (8)$$

Results about item discrimination: Ind and Min models



The area between the Min and the Ind IRFs along the line $\theta_1 = \theta_2$ is $1/\alpha$.
Moreover, this is the line along which the area is largest.

Results about item discrimination: AI and Max models

Letting $\zeta = \max\{\theta_j, \theta_k\}$, then

$$P_i(\theta_j) + P_i(\theta_k) - P_i(\theta_j) P_i(\theta_k) \leq 2P_i(\zeta) - P_i(\zeta)^2, \quad (9)$$

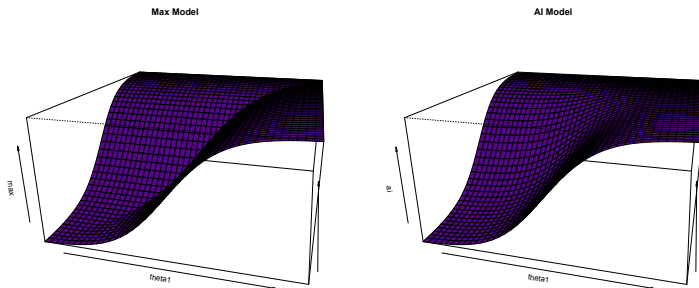
an upper bound on the difference between AI IRF and the max IRF is

$$[2P_i(\zeta) - P_i(\zeta)^2] - P_i(\zeta) = \Delta_i(\zeta), \quad (10)$$

leading to

$$\int_{-\infty}^{\infty} \Delta_i(\zeta) d\zeta = \frac{1}{\alpha_i}. \quad (11)$$

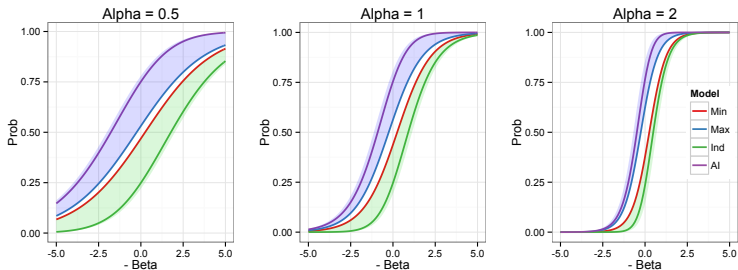
Results about item discrimination: AI and Max models



The area between the AI and the Max IRFs along the line $\theta_1 = \theta_2$ is $1/\alpha$. Moreover, this is the line along which the area is largest.

Results about item discrimination: Example “CRFs”

Illustration of differences among the Independence, Min, Max, and AI models, as a function of item discrimination for collaborators with $\xi = -.25$ and $\zeta = .25$. The blue area denotes $\Delta(\zeta)$ and the green area denotes $\Delta(\xi)$.



Results about abilities: Ind and Min models

- ▶ Letting $\zeta = \xi + u$ for $u \in \mathbb{R}_+$ and assume $\beta_i = 0$, the upper limit in u of the Ind model just the Min model:

$$\lim_{u \rightarrow \infty} P_i(\xi + u)P_i(\xi) = P_i(\xi). \quad (12)$$

The performance of the two independent individuals becomes increasingly similar to that of the less able person, when the item is not challenging for the more able person.

Results about abilities: AI and Max models

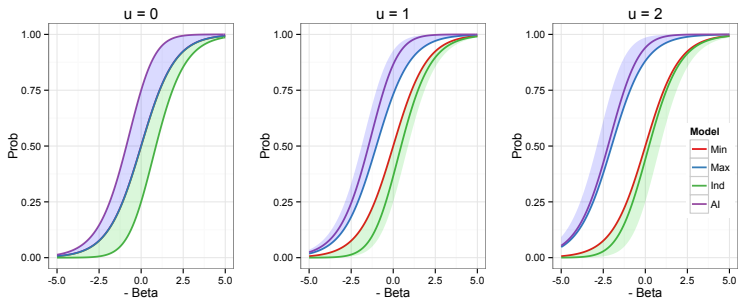
- The upper limit in u of the difference between AI model and the Max model:

$$\lim_{u \rightarrow \infty} P_i(\xi)(1 - P_i(\xi + u)) = 0. \quad (13)$$

When one individual is very likely to answer an item correctly, the contribution of the less able person becomes increasingly negligible.

Results about abilities: example CRFs

Illustration of differences among the Independence, Min, Max, and AI models, as a function of u for collaborators with $\zeta = 0$ and $\xi = \zeta + u$ and item parameters $\alpha = 1$ and $\beta = 0$. The green area denotes $\Delta(\zeta)$ and the blue area denotes $\Delta(\xi)$.



Part 2: denouement

- ▶ An assessment designed to provide evidence about the presence or absence of any additive model of successful collaboration¹⁰ should involve:
 - a) Collaborators with relatively similar abilities
 - b) Answering items with relatively low discrimination that are targeted at their ability level
- ▶ $1/\alpha$ is useful to gauge the item-by-item distinguishability of the Max and AI models when $\theta_j \approx \theta_k$

¹⁰ Recall that the AI model is an upperbound on any additive model of successful collaboration – slide 25

Part 3: One model to rule them all!

Let $w_1, w_2 \in [0, 1]$ and define the weighted additive independence model

$$E_{\text{WAI}}[X_{ijk} \mid \theta_j \theta_k] = w_j P_i(\theta_j) Q_i(\theta_k) + w_k P_i(\theta_k) Q_i(\theta_j) + P_i(\theta_j) P_i(\theta_k)$$

- ▶ Includes original four and everything in between
- ▶ Includes $(P_i(\theta_j) + P_i(\theta_k))/2$ when $w_1 = w_2 = .5$
- ▶ Weights describe how well each individual obtains his/her “optimal collaboration level”

Part 3: The weighted AI model

But estimation problems abound:

- ▶ Original four are on boundary of parameter space
- ▶ Likelihood includes weighted sum of density-like terms (yuck)
- ▶ Item and test information for weights are either hard or bad, depending on approach

Current research questions

- ▶ Is it worth pursuing the WAI model, despite its problems?
- ▶ Are there other technical / interpretative difficulties with this approach?
- ▶ Are there better options for modeling collaborative outcomes?

Contact: peter.halpin@nyu.edu

Collaborators: Yoav Bergner, ETS; Jacqueline Gutman, NYU

Support: This research was funded by a postdoctoral fellowship from the Spencer Foundation and an Education Technology grant from NYU Steinhardt.