Workshop 1: Item design for assessments involving collaboration

Peter F. Halpin

April 18, 2016

`github.com/peterhalpin/BearShare`

NEW YORK UNIVERSITY

Outline

Part 1: Wherefore assessments involving collaboration?

- Set up the current perspective: performance assessments
- Uses of collaboration in assessment contexts

Outline

Part 1: Wherefore assessments involving collaboration?

- Set up the current perspective: performance assessments
- Uses of collaboration in assessment contexts

Part 2: Item design for collaboration

- What's wrong with using "one-player" items for collaboration?
- Some examples of two-player items obtained from one-player items

Outline

Part 1: Wherefore assessments involving collaboration?

- Set up the current perspective: performance assessments
- Uses of collaboration in assessment contexts

Part 2: Item design for collaboration

- What's wrong with using "one-player" items for collaboration?
- Some examples of two-player items obtained from one-player items

Part 3: OpenEdx and CPSX for building your own collaborative assessments

- Git repo: `github.com/ybergner/cpsx.git`
- Live version: `collaborative-assessment.org`

Part 1: Why?

- 21st-century skills, non-cognitive skills, soft skills, hard-to-measure skills, social skills, ...
  - Theme: traditional educational tests target a relatively narrow set of constructs

Part 1: Why?

- 21st-century skills, non-cognitive skills, soft skills, hard-to-measure skills, social skills, ...
  - Theme: traditional educational tests target a relatively narrow set of constructs
- Analyses of US labour markets indicate that such skills are valued by employers (Burrus et al. 2013, Deming 2015)

Part 1: Why?

- 21st-century skills, non-cognitive skills, soft skills, hard-to-measure skills, social skills, ...
  - Theme: traditional educational tests target a relatively narrow set of constructs
- Analyses of US labour markets indicate that such skills are valued by employers (Burrus et al. 2013, Deming 2015)
- There is a salient demand for assessments of a broader range of student competencies

With apologies to Dr. Duckworth...

8- Item Grit Scale

Directions for taking the Grit Scale: Please respond to the following 8 items. Be honest – there are no right or wrong answers!

1. New ideas and projects sometimes distract me from previous ones.*
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

2. Setbacks (delays and obstacles) don't discourage me. I bounce back from disappointments faster than most people.
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

3. I have been obsessed with a certain idea or project for a short time but later lost interest.*
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

`upenn.app.box.com/8itemgrit`

Self-reports

- Self-report measures often do not require the respondent to exhibit the skills about which we wish to make inferences

  → Unsuitable for supporting consequential decisions in educational settings[1]

[1] cf. Duckworth, & Yeager (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44(4)*, 237-251.

Educational assessments

- ☺ Reliability and generalizability in traditional content domains

[2] www.wnyc.org/story/new-york-city-students-make-modest-gains-state-tests-opt-out-numbers-triple/

Educational assessments

- ☺ Reliability and generalizability in traditional content domains
- ☹ Current psychometric models don't seem entirely appropriate to "next generation assessments"
  - e.g., IRT models don't use process data

[2] www.wnyc.org/story/new-york-city-students-make-modest-gains-state-tests-opt-out-numbers-triple/

Educational assessments

- ☺ Reliability and generalizability in traditional content domains
- ☹ Current psychometric models don't seem entirely appropriate to "next generation assessments"
  - e.g., IRT models don't use process data
- ☹ Collateral damage: teaching to the test, test anxiety, bubble-filling, ...
  - NY opt-out movement: 20% of students (parents) boycotted state test last year[2]

[2] www.wnyc.org/story/new-york-city-students-make-modest-gains-state-tests-opt-out-numbers-triple/

Performance assessments[3]

A performance assessment (sometimes called a work sample when assessing job performance), as defined in this report, is an activity or set of activities that requires test takers, either individually or in groups, to generate products or performances in response to a complex, most often real-world task. These products and performances provide observable evidence bearing on test takers' knowledge, skills, and abilities—their competencies—in completing the assessment (e.g., Shavelson, 2013). Such assessments as science performance assessments, essays using informative documents, portfolios, computer simulations, projects, and demonstrations may be considered forms of performance assessment.

[3] Davey, Ferrara, Holland, Shavelson, Webb, & Wise (2015). Psychometric Considerations for the Next Generation of Performance Assessment. Princeton, NJ. p. 10

Collaboration as a modality of performance assessment

- Small group interactions are a highly-valued educational practice
  - The Jigsaw Classroom (Aronson et al. 1978; `jigsaw.org`)
  - Group-worthy tasks (Cohen et al. 1999)
- The use of information technology to support student collaboration is well established
  - CSCL (e.g., Hmelo-Silver et al. 2013)

Collaboration as a modality of performance assessment

- Small group interactions are a highly-valued educational practice
  - The Jigsaw Classroom (Aronson et al. 1978; `jigsaw.org`)
  - Group-worthy tasks (Cohen et al. 1999)
- The use of information technology to support student collaboration is well established
  - CSCL (e.g., Hmelo-Silver et al. 2013)
- The use of group work in assessment contexts has a relatively long-standing history
  - Webb 1995, 2014

Webb 1995[4]

Because the processes and outcomes of group collaboration may differ depending on whether the goal is individual learning or group productivity, it is important that the purpose of the assessment, the goal of group work, and the group processes supposed to contribute to those goals be specified clearly.

1. Purpose of the assessment
2. The goal of the group work
3. Processes of the group work

[4] Group Collaboration in Assessment: Multiple Objectives, Processes, and Outcomes. *Educational Evaluation and Policy Analysis, 17(2)*, p. 241

Webb's three-part theory of collaboration in assessments

Purpose of assessment

1. Individual learning after collaboration
2. Group productivity
3. Students' ability to work together

Goals of group work

1. Individual learning
2. Group productivity

Group processes

1. Co-construction of ideas
2. Giving and receiving help
3. Conflict and cooperation
4. Equality of participation
5. Social loafing
6. Division of labour

Webb's three-part theory of collaboration in assessments

Purpose of assessment

1. Individual learning after collaboration
2. Group productivity
3. Students' ability to work together

Goals of group work

1. Individual learning
2. Group productivity

Group processes

1. Co-construction of ideas
2. Giving and receiving help
3. Conflict and cooperation
4. Equality of participation
5. Social loafing
6. Division of labour

Webb's three-part theory of collaboration in assessments

Purpose of assessment

1. Individual learning after collaboration
2. Group productivity
3. Students' ability to work together

Goals of group work

1. Individual learning
2. Group productivity

Group processes

1. Co-construction of ideas
2. Giving and receiving help
3. Conflict and cooperation
4. Equality of participation
5. Social loafing
6. Division of labour

Webb's three-part theory of collaboration in assessments

Purpose of assessment

1. Individual learning after collaboration
2. Group productivity
3. Students' ability to work together

Goals of group work

1. Individual learning
2. Group productivity

> For tasks with a correct answer or single best solution, then, allowing the most competent individual or individuals to dominate the group may be the most efficient and effective strategy to maximize group productivity.

Group processes

1. Co-construction of ideas
2. Giving and receiving help
3. Conflict and cooperation
4. Equality of participation
5. Social loafing
6. Division of labour

Webb's three-part theory of collaboration in assessments

Purpose of assessment

1. Individual learning after collaboration
2. Group productivity
3. Students' ability to work together

Goals of group work

1. Individual learning
2. Group productivity

When the task is nonroutine, does not have well-specified procedures, and requires input from all group members, equal participation will be important for both individual learning and group productivity, making these goals complementary for all group members.

Group processes

1. Co-construction of ideas
2. Giving and receiving help
3. Conflict and cooperation
4. Equality of participation
5. Social loafing
6. Division of labour

Webb's three-part theory of collaboration in assessments

Purpose of assessment

1. Individual learning after collaboration
2. Group productivity
3. Students' ability to work together

Goals of group work

1. Individual learning
2. Group productivity

When the task is nonroutine, does not have well-specified procedures, and requires input from all group members, equal participation will be important for both individual learning and group productivity, making these goals complementary for all group members.

Group processes

1. Co-construction of ideas
2. Giving and receiving help
3. Conflict and cooperation
4. Equality of participation
5. Social loafing
6. Division of labour

Is this possible?

Purpose of assessment

1. Individual learning after collaboration
2. Group productivity
3. Students' ability to work together

Goals of group work

1. Individual learning
2. Group productivity

Group processes

1. Co-construction of ideas
2. Giving and receiving help
3. Conflict and cooperation
4. Equality of participation
5. Social loafing
6. Division of labour

Part 2: Item design

[5] Organisation for Economic Co-operation and Development. (2013). PISA 2015 Draft Collaborative Problem Solving Framework. Retrieved from http://www.oecd.org/pisa/pisaproducts/DraftPISA2015CollaborativeProblemSolvingFramework.pdf

Part 2: Item design

- PISA 2015 CPS[5]
  - Various item types: jigsaw / information sharing; consensus building; negotiation
  - Problems were interactive, but with simulated collaboration (deterministic computer agent)
  - Problems were not designed to assess content knowledge in a traditional domain – "below grade level"

[5] Organisation for Economic Co-operation and Development. (2013). PISA 2015 Draft Collaborative Problem Solving Framework. Retrieved from http://www.oecd.org/pisa/pisaproducts/DraftPISA2015CollaborativeProblemSolvingFramework.pdf

Item design

- ATC21S[6]
  - 4 prototype tasks[7] (3 similar to what I will discuss)
  - Problems were interactive, and with interactions between real students
  - Problems were designed for learning as well as assessment in target domains

[6] Griffin & Care (2015). Assessment and teaching of 21st century skills: Methods and approach

[7] http://www.atc21s.org/uploads/3/7/0/0/37007163/pd_module_3_nonadmin.pdf

Item design

- ATC21S[6]
  - 4 prototype tasks[7] (3 similar to what I will discuss)
  - Problems were interactive, and with interactions between real students
  - Problems were designed for learning as well as assessment in target domains
- This doesn't really let YOU design tasks for collaboration

[6] Griffin & Care (2015). Assessment and teaching of 21st century skills: Methods and approach

[7] `http://www.atc21s.org/uploads/3/7/0/0/37007163/pd_module_3_nonadmin.pdf`

Item design

- DIY with CPSX[8]
  - Built on OpenEdx
    - Open source LMS, easy to set up (locally or AWS), active development
  - 40+ built-in problem types with automatic grading and adaptive hints[9]
    - Traditional (MC, NR); interactive; customizable (`xml, JS`)
  - CPSX allows for small group chat during problem solving
    - Currently in its infancy – e.g., no screen sharing / shared manipulables

[8] `https://github.com/ybergner/cpsx.git`
[9] `http://docs.edx.org`

Goals of the current approach

1. Make use of tech that's already available and free (speech and beer)

2. Tasks clearly anchored in an educational content domain
   - Specifically math

3. Navigate "minimal design" vs group-worthy tasks
   - Working from psychometric theory towards best practices in group work, rather than the other way around

Modifying one-player items

- How can a conventional mathematics test question be adapted to a collaborative context?
  - Recipes for creating collaborative tasks from existing assessment materials
- Not an ideal approach to designing group-worthy tasks
- However, retain the strengths of existing assessments
  - e.g., Calibrate the one-player version to anchor the domain difficulty of the two-player version

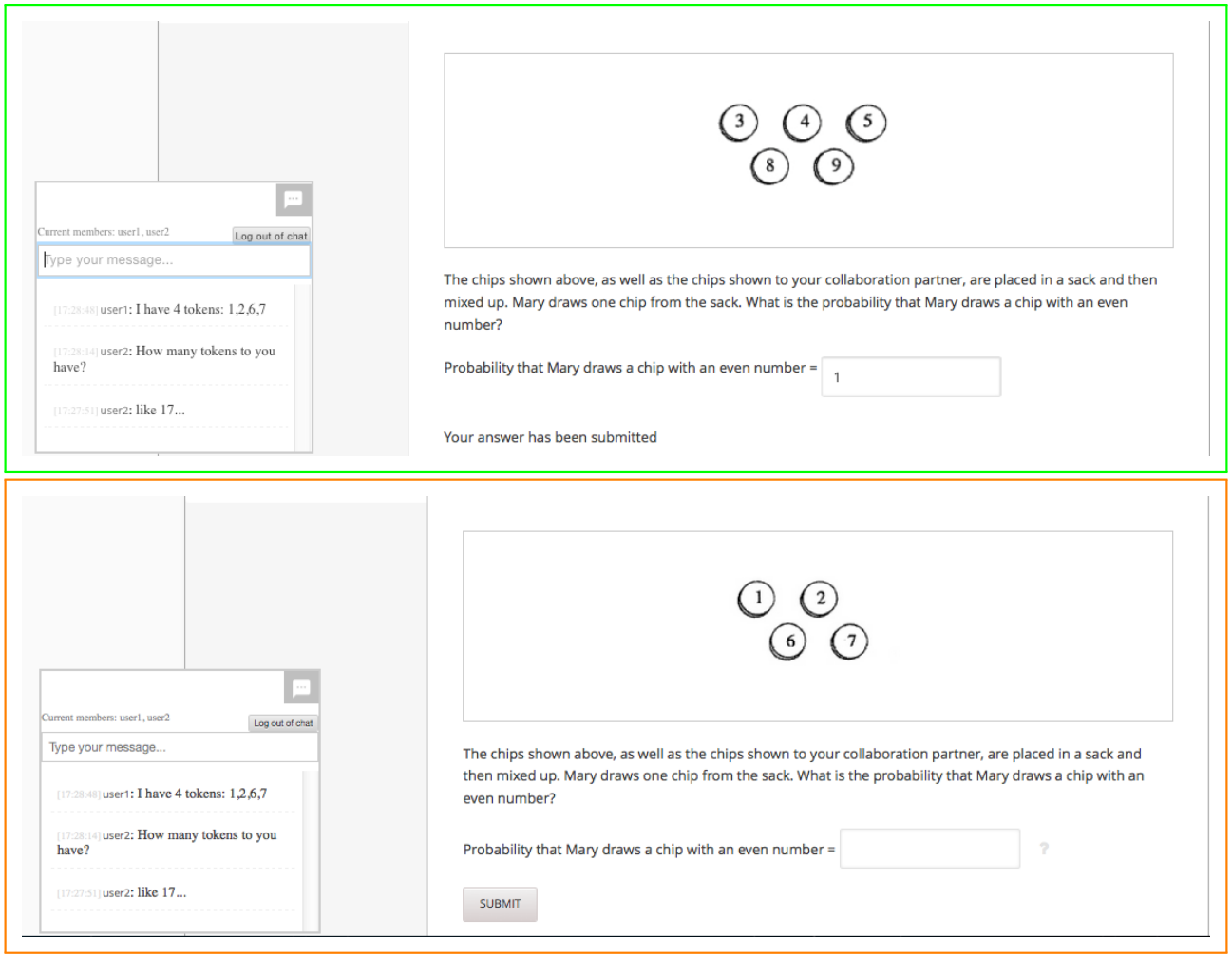
Standard items

- Change in the instructions while retaining the assessment materials
  - e.g., two students, one copy of a math test, evaluation depends only on what they record on the test form

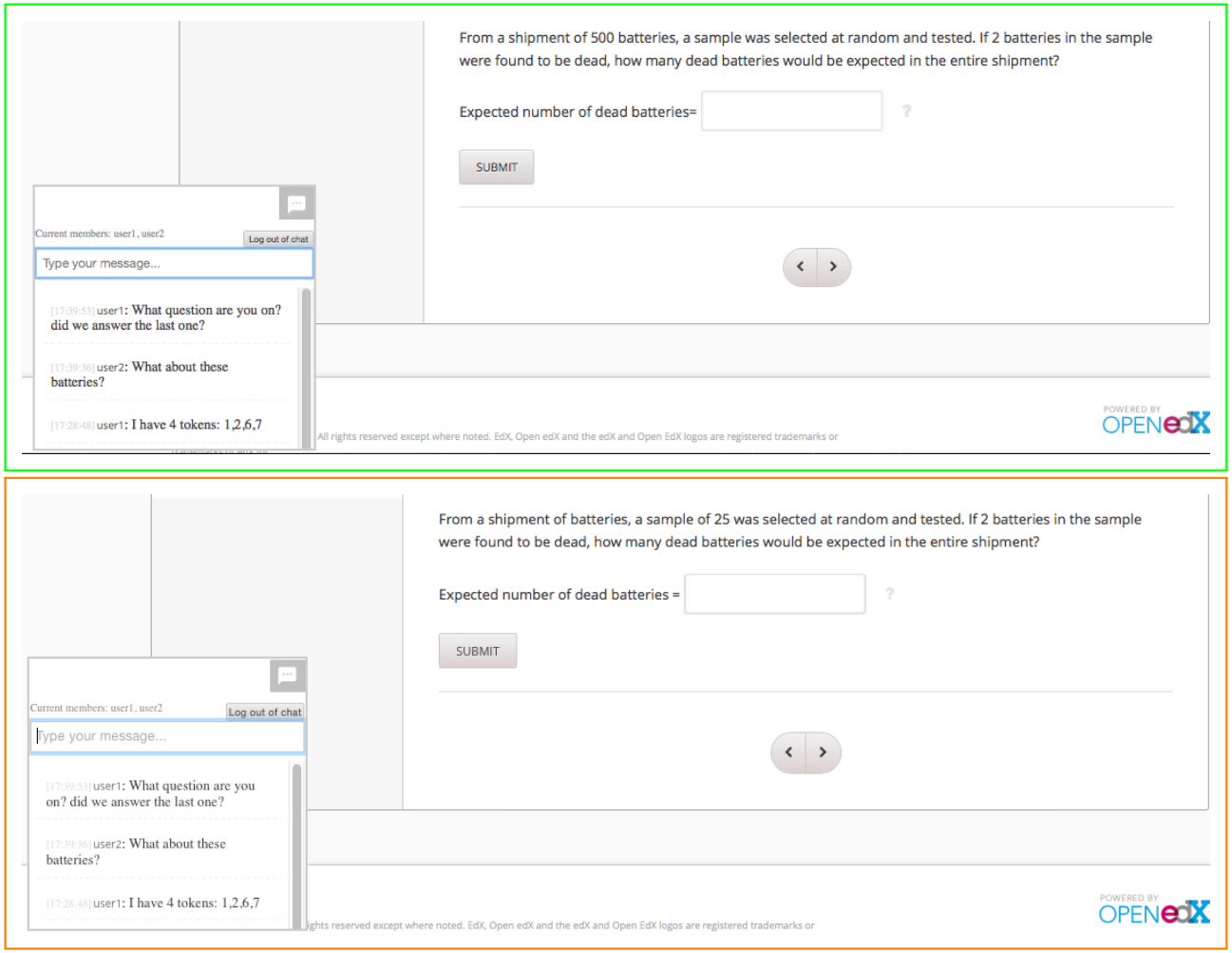
☺ Minimal design

☹ Group-worthy tasks

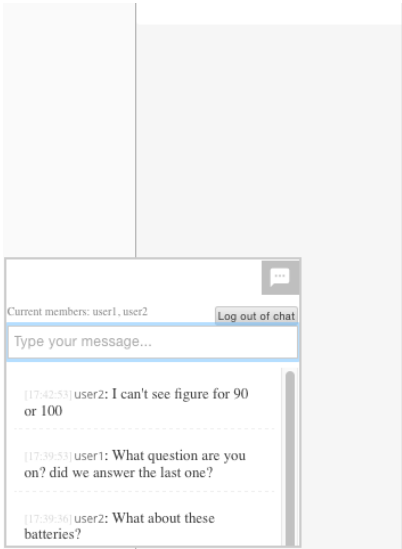
Jigsaw / information sharing items



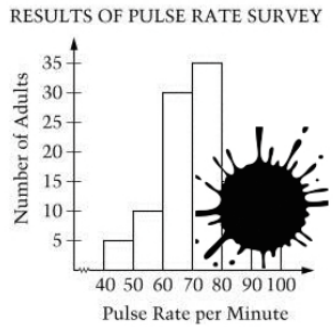
Jigsaw / information sharing items



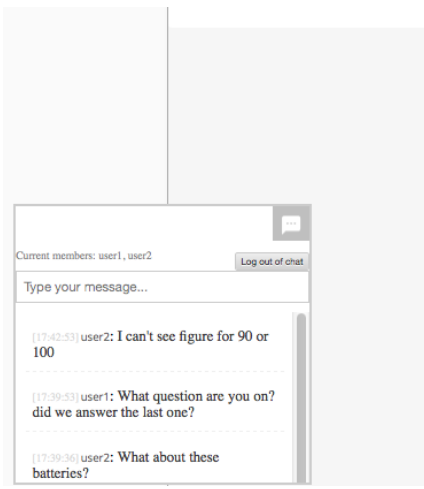
Jigsaw / information sharing items



The following question refers to the data shown below, as well as the data shown to your collaboration partner. Your figure shows the pulse rate per minute of a group of 100 adults. For example, 5 adults have a pulse rate from 40-49 inclusive. However, some ink has been spilled on the figure.



What is the average pulse rate per minute for these 100 people?



The following question refers to the data shown below, as well as the data shown to your collaboration partner. Your table shows the pulse rate per minute of a group of 100 adults. For example, 5 adults have a pulse rate from 40-49 inclusive. However, some ink has been spilled on the table.

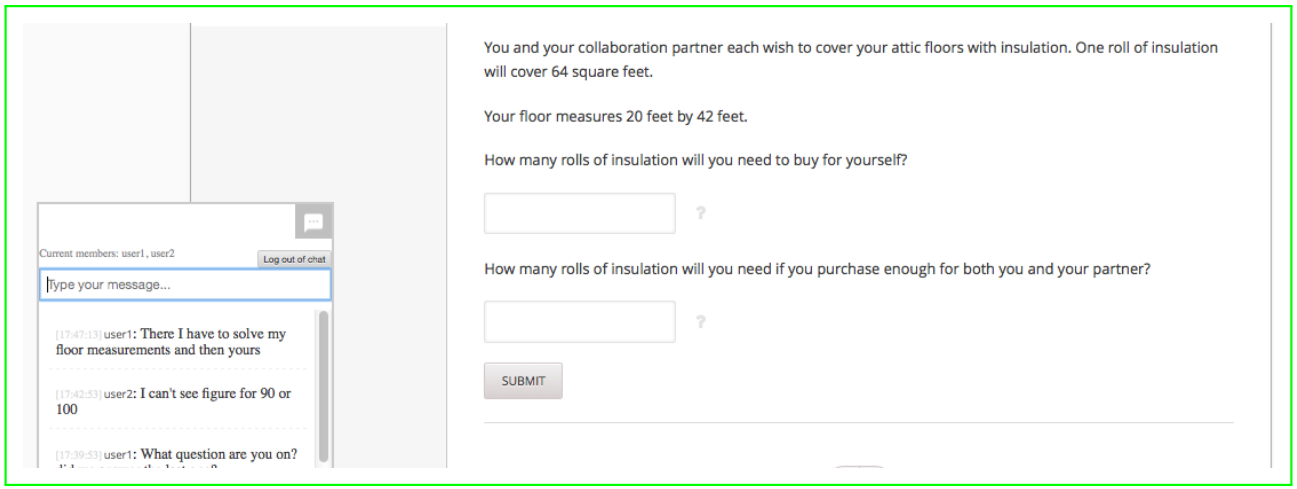
PULSE RATE PER MINUTE	NUMBER OF ADULTS
40-49	5
50-59	
60-69	30
70-79	
80-89	15
90-99	5

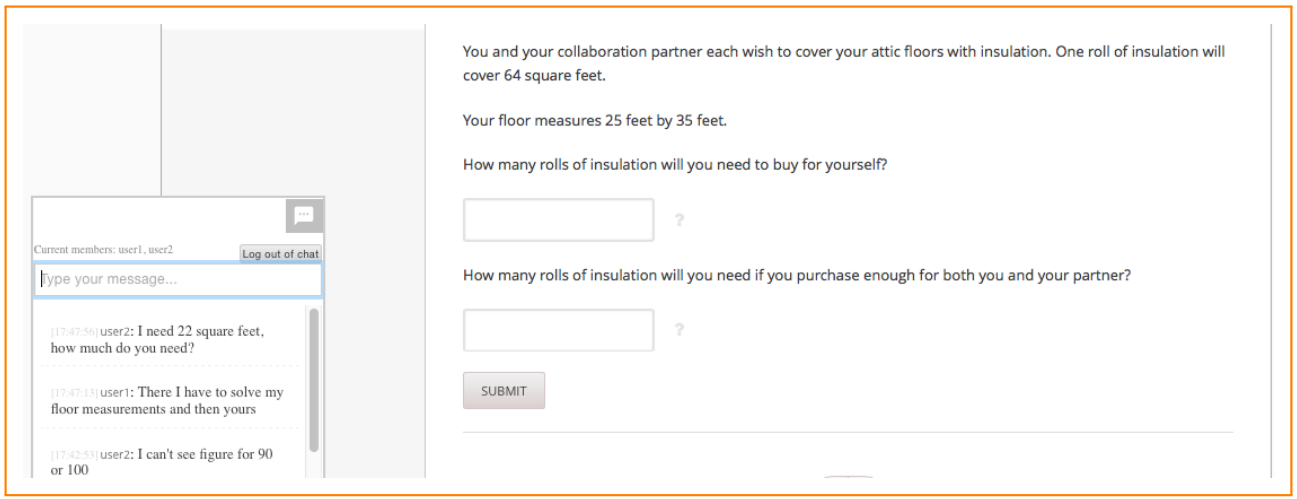
What is the average pulse rate per minute for these 100 people?

(NOTE: Use the midpoint of each interval to represent the pulse rate for the entire interval. For example, 55 would be used for the pulse rate of the people in the 50-59 group.)

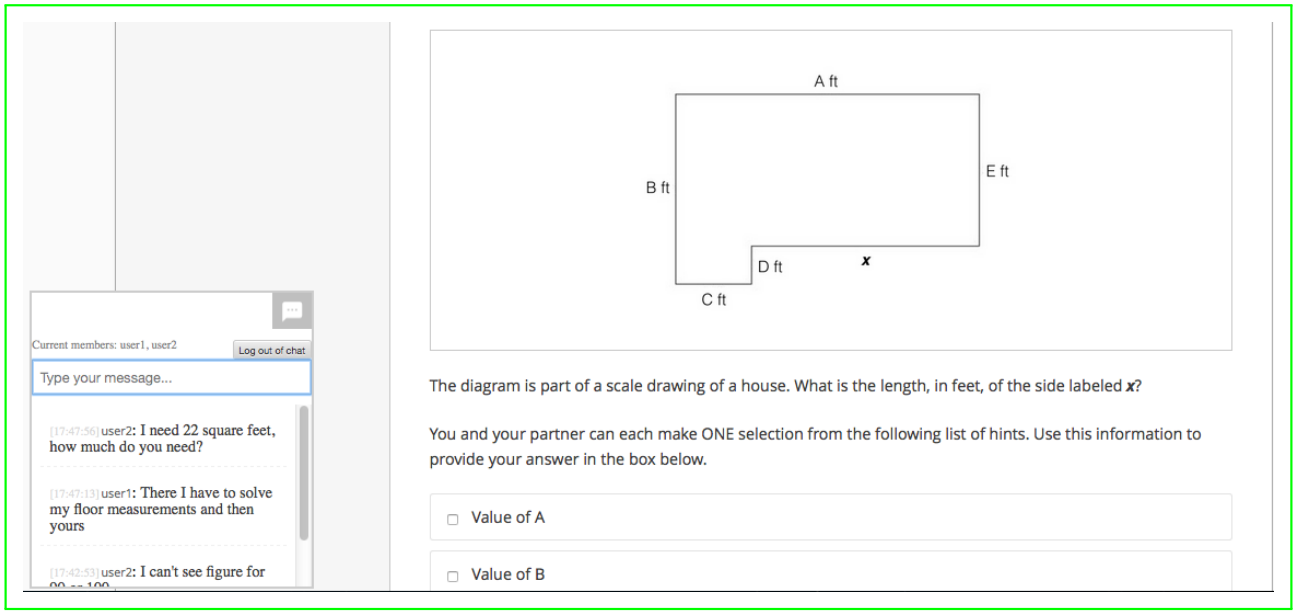
Average pulse rate =

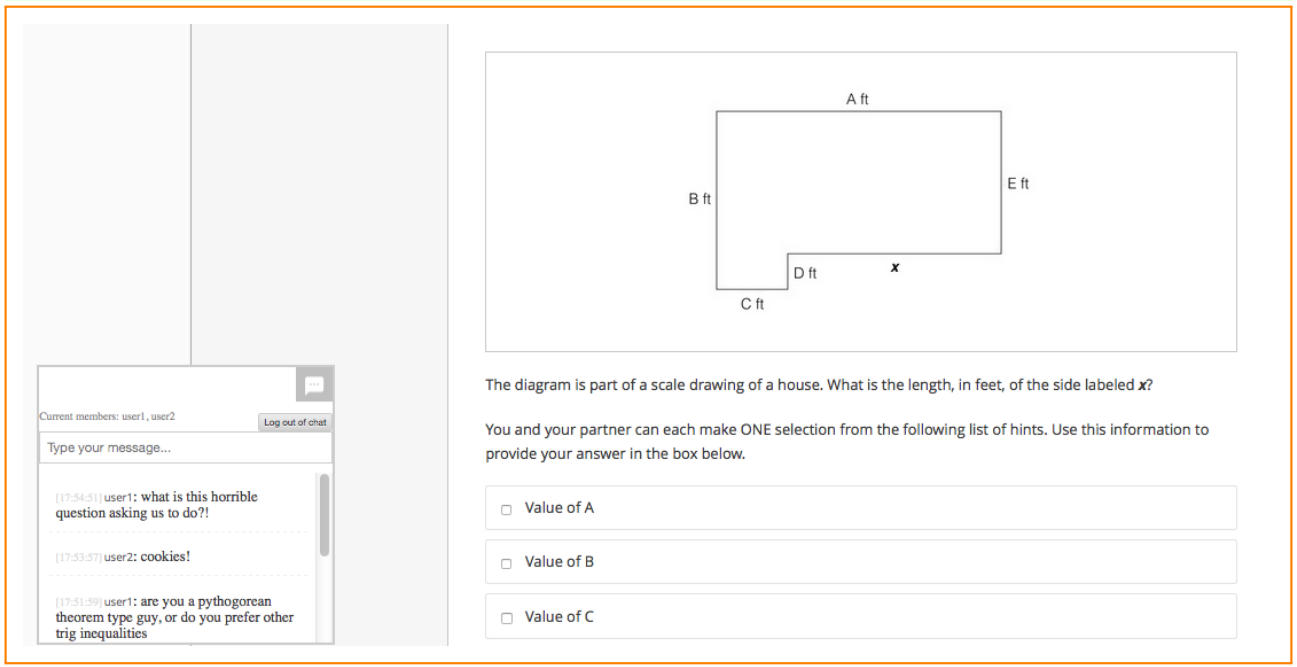
Jigsaw / information sharing items



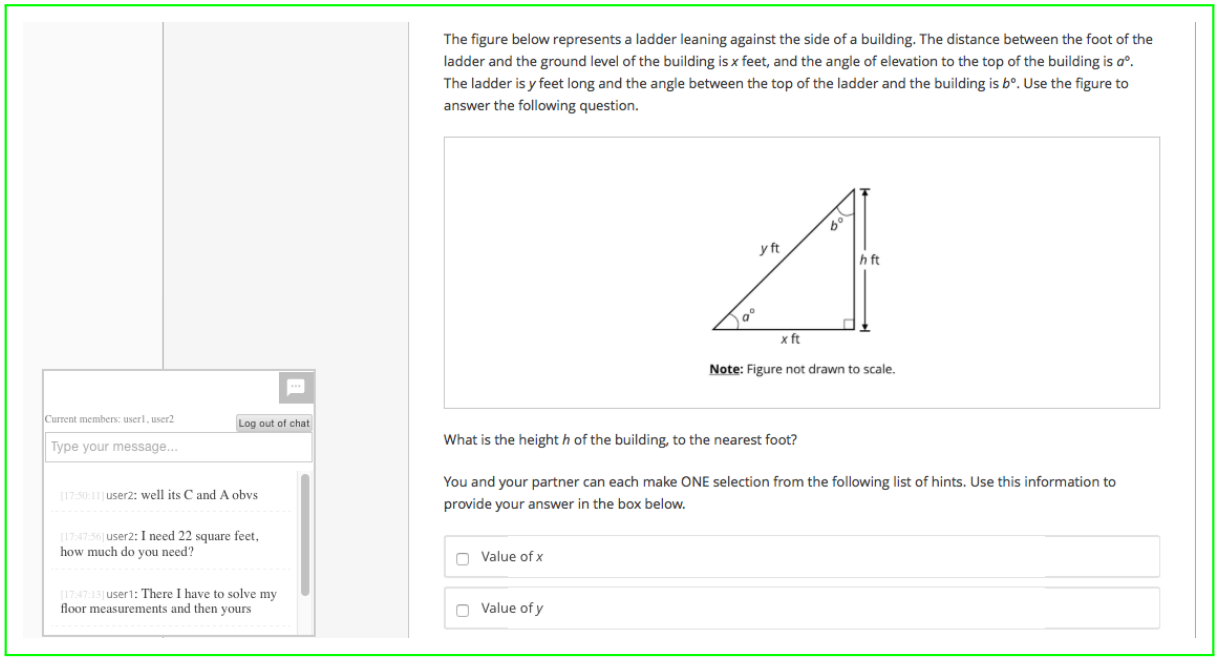


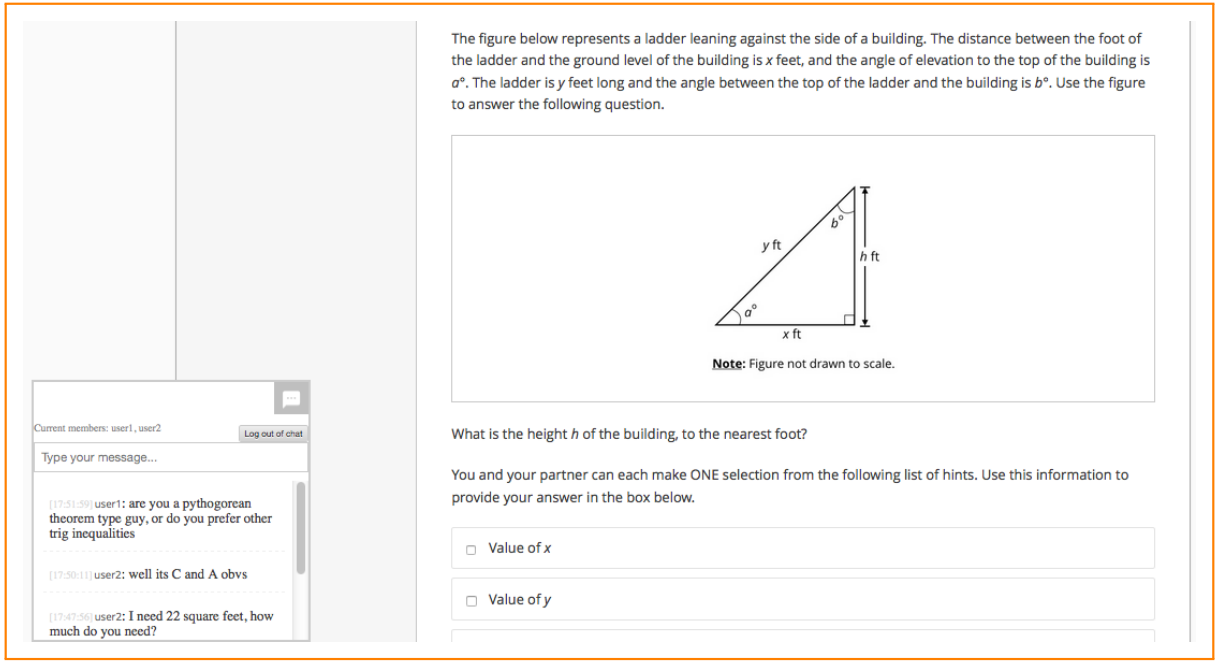
Hints / information requesting items



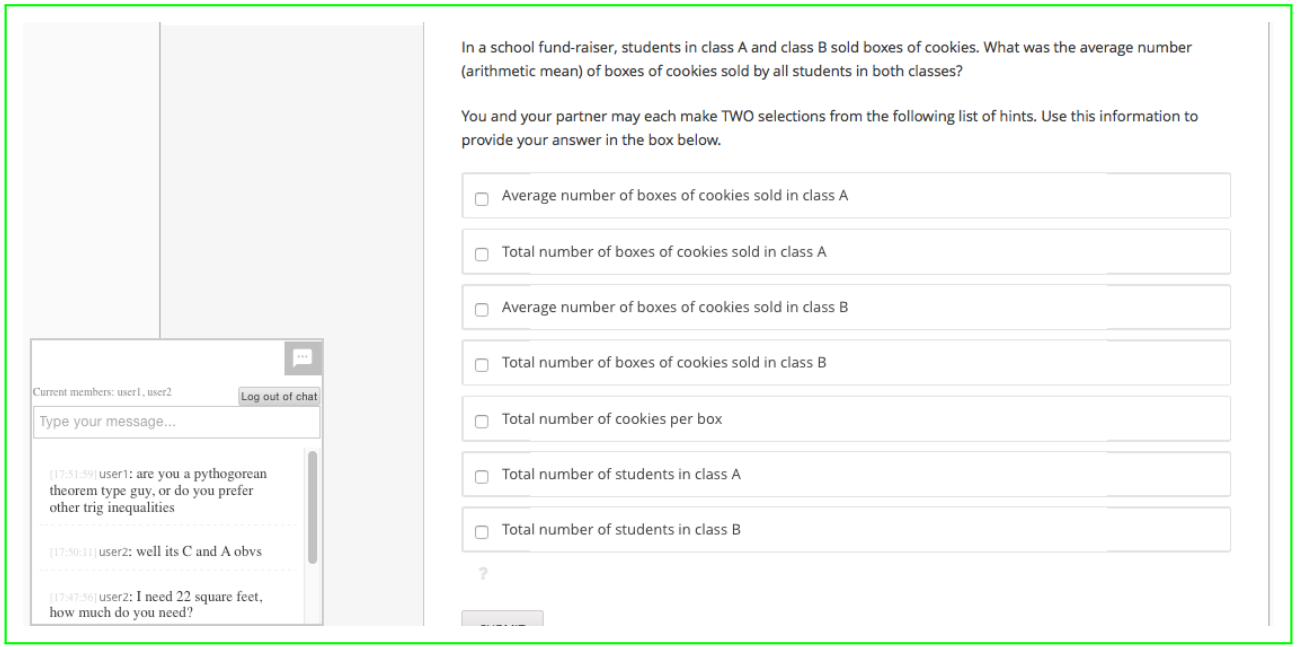


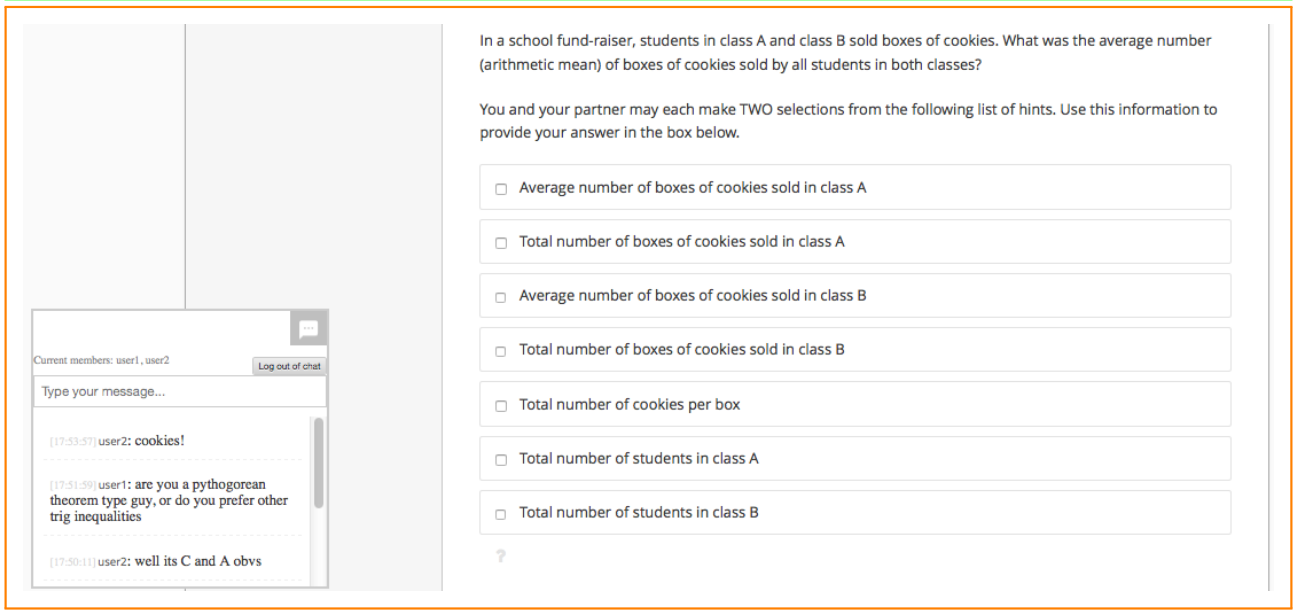
Hints / information requesting items



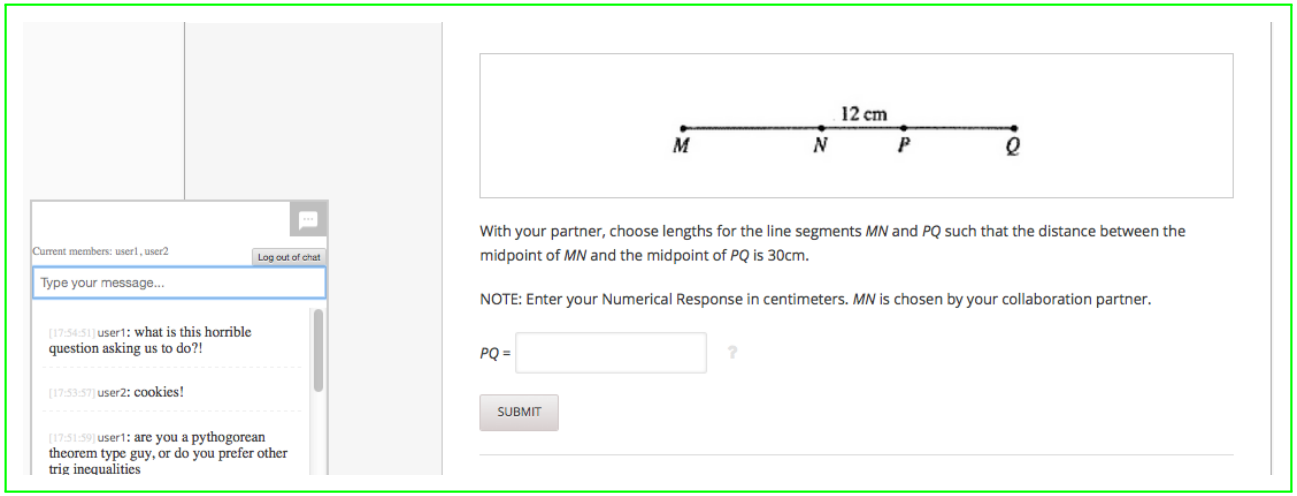


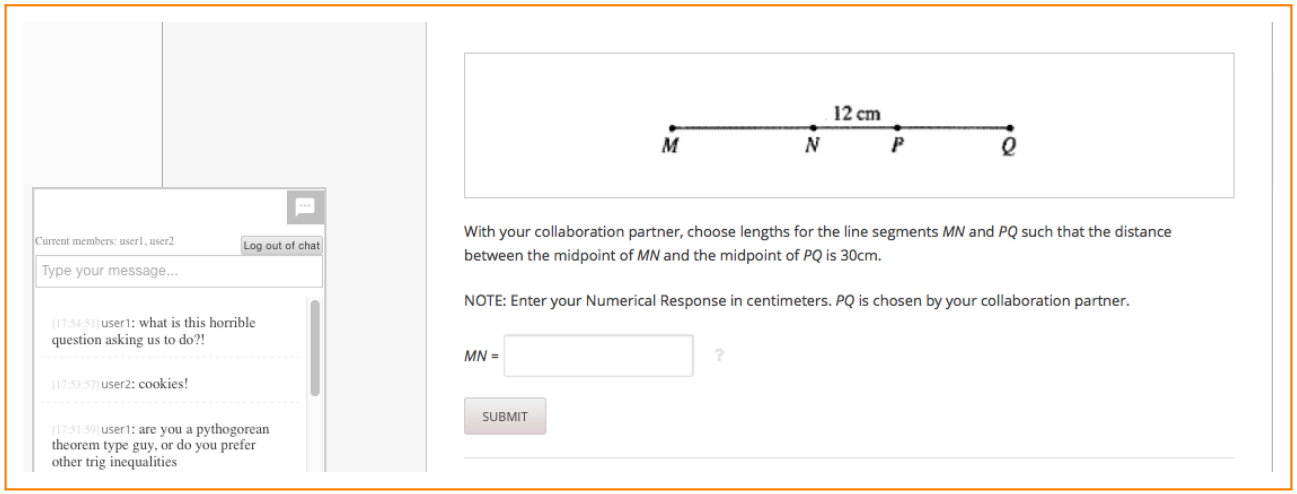
Hints / information requesting items





Multiple answer / negotiation items





A note on process loss[10]

- Process loss as the discrepancy between potential and actual performance
  - Actual Productivity = Potential Productivity - Process Loss
- Two main sources: motivation and coordination, often treated as functions of group size
- For our two-player items, process loss due to coordination can be thought of as an item characteristic

[10] Steiner (1972) Group processes and productivity

A IRF modified for process loss

- Assume 2PL for one-player items i

$$\text{logit}(p_{ij}) = \alpha_i \, \theta_j + \beta_i$$

- Process loss for two-player version i':

$$\text{logit}(p_{i'j}) = \alpha_i \, \theta_j + \beta_i + \gamma_{i'}$$

Viable designs for estimating process loss

- Independent samples or repeated measures with counterbalanced forms where:
  - An individual assessment serves as the calibration sample for item i
  - A collaborative assessment serves as the calibration sample for item i'
  - With $\sum_{i'} \gamma_{i'} = 0$ to identify $E[\theta_j]$ in the collaborative sample

Viable designs for estimating process loss

- Independent samples or repeated measures with counterbalanced forms where:
  - An individual assessment serves as the calibration sample for item i
  - A collaborative assessment serves as the calibration sample for item i'
  - With $\sum_{i'} \gamma_{i'} = 0$ to identify $E[\theta_j]$ in the collaborative sample

- But what is θ_j in the collaborative sample? More on that tomorrow!

OpenEdx / CPSX demo

- LMS (Students): `collaborative-assessment.org`
- CMS (Content developer): `collaborative-assessment.org:18010`
- Login credentials: For (i in $100:120$)
  - email: user$i$@example.com
  - password: user$i$

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.

Aronson, E., Blaney, N., Stephan, C., Sikes, J., Snapp, M. (1978). The jigsaw classroom. Beverly Hills, CA: Sage.

Cohen, E. G., Lotan, R. A., Scarloss, B. A., & Arellano, A. R. (1999). Complex instruction: Equity in cooperative learning classrooms. Theory Into Practice, 38, 80-86.

Davey, T., Ferrara, S., Holland, P. W., Shavelson, R. J., Webb, N. M., Wise, L. L. (2015). Psychometric Considerations for the Next Generation of Performance Assessment. Princeton, NJ.

Deming, D. J. (2015). The Growing Importance of Social Skills in the Labor Market. National Bureau of Economic Research Working Paper Series, (21473).

Griffin, P., Care, E. (2015). Assessment and teaching of 21st century skills: Methods and approach. New York: Springer.

Hmelo-Silver, C. E., Chinn, C. A., Chan, C. K., & O?Donnel, A. M. (2013). International handbook of collaborative learning. New York: Taylor and Francis.

Organisation for Economic Co-operation and Development. (2013). PISA 2015 Draft Collaborative Problem Solving Framework. Retrieved from http://www.oecd.org/pisa/pisaproducts/DraftPISA2015CollaborativeProblemSolvingFramework.pdf

Webb, N. M. (1995). Group Collaboration in Assessment: Multiple Objectives, Processes, and Outcomes. Educational Evaluation and Policy Analysis, 17(2), 239-261.