Computer Game Design for Affective Assessment

Harry O'Neil

University of Southern California/CRESST

CRESSTCON
Los Angeles, CA
October 1-2, 2018

Affective Assessment

- Cognitive, Affective, Psychomotor
  - Thinking vs. Feeling vs. Doing
- Affective Constructs
  - Also similar constructs: socio-emotional, 21st Century Skills, non-cognitive, competencies, soft skills
  - Desirable in their own right
    - PISA
  - Related to academic achievement
    - Effect sizes of .3 to .4 in the literature (Hattie, 2008)

Affective Assessment Measures

- Learning analytics (Loh, Sheng, and Ifenthaler, 2015) are viewed as a focus on real-time learning processes based on educational information in games and simulations
    - Both affective and cognitive
    - It is used for real-time interpretation, modeling, and prediction
- Metrics the learners' individual characteristics, such as socio-demographic information, personal preferences, interests, responses to inventories (e.g., personality), skills and competencies, prior knowledge and academic performance

Affective Assessment Issues

- Traits vs. States
  - Traits are considered relatively enduring predispositions or characteristics of people (e.g., aptitude)
  - States are attributes of individuals' variability over time or occasions (Spielberger, 1975)
    - Manifestations of state characteristics are highly dependent on the environment and circumstances of the specific instance
- Generalized vs. Situational Specific Contexts
- Purposes
  - The use of Trait measures for selection purposes and State measures for program evaluation, diagnostic accountability and formative assessment purposes.

Mayer's (2015) Three Genres of Game Research

- Media comparison research (Do people learn academic content better from games than from conventional media?)
- Value Added research (Does adding feature X to a game improve learning?)
- Cognitive and Affective consequences research (What do people learn from playing a game?)

Affective Consequences

- How are State vs. Trait affective variables measured via self report?
  - Instructions: Tell me how you felt during the game vs. tell me your feelings in general
  - Ratings of intensity vs. frequency
  - Different items
- What is state efficacy test anxiety, interest and state effort?

Game Context

- What was our game?
  - PuppyBot designed by CMU
  - Teaches middle school physics to elementary school students
  - Achievement and affective measures
- What is the process for affective measures?
  - Usability study for ratings via Amazon Turk with adults
    - If adults cannot do the task then kids can not. However, if adults can do the task, kids may not be able to do the tasks
  - Usability study via face-to-face with kids
  - Revision of items

Summary of Usability Results

- Feasibility of using Amazon Turk for formative evaluation of games icons to measure affective states
- What icons work best for adults?
  - Intensity, circles (small, medium, large)
- Do they think they will work for kids?
  - Yes

Descriptive Statistics

	Effect Size	Alpha Reliability
Self Efficacy (5 items)	0.40	0.68
Effort (6 items)	0.48	0.80
Post-test (20 items)	0.23	0.84

Summary

- PuppyBot has affective consequences
  - Both state self-efficacy and effort were significantly higher compared to control game
    - Effect sizes were moderate
      - Experimental group was at the 65 to 70 percentile with control group at the 50th percentile
- New measures of state self-efficacy, and effort with adequate reliability for children
  - Some validity information

Interest Construct

Context	Construct	Definition
Navy Job	Work Interest	“Relatively stable individual differences, grounded in an individual's identity, that encompass one's preferences for performing selected work activities or working in certain environments (or contexts) that purposefully influence work-related choices and behavior (e.g., occupational or job choice) through motivational processes" (p. 166 Ingerick & Rumsey, 2014)
Game	Situational Interest	”Focused attention and the affective reaction that is triggered in the moment by environmental stimuli, which may or may not last over time" (p.113, Hidi & Renninger, 2006)

Affective Assessment Measures (Cont'd)

- State Self-efficacy, Interest and Test Anxiety
  - Measurement of self-efficacy and test anxiety (state worry) by five item self-report scales (O'Neil et al., 19xx)
  - We measure interest two ways:
    - Job Opportunities in the Navy metric (JOIN)
    - The US Navy has embraced using interest as predictor of occupational success. It has demonstrated good reliability (.88) (Farmer et al., 2003)
    - Game-based measures

Game-based Measures of Interest

Choice Measures	• Choice in playing an additional level • Choice in playing game even when not required (having the game available to participant even after study is over) • Choice in playing a more difficult level • Choice in playing a rating-related level versus control level • Exploration of Navy ship (going to areas not necessarily prescribed in the game) • Time spent with seductive details (elements that are inserted into the game that are not relevant to game-play) • Time spent with seductive details of another rating (might suggest disinterest in the current rating)
In-game Experience Measures	• Experience sample method of prompting players to report affect while playing the game, e.g., self-efficacy
Information Seeking	• Player has the option of seeking additional information about the rating • Player has the option of seeking additional information about a different rating (might suggest disinterest in the current rating)

Ongoing Research: Affective Measurement

- How are self-efficacy, interest, and test anxiety measured via game learning analytics and metrics?
- JOIN Navy Life Game
  - Ability, interest and other affective measures (e.g., self-efficacy), for Navy recruitment
  - 1% reduction in attrition cost-avoidance of 8 million
- NETC Assessment Framework
- “Assessment Issues in Simulation and Games” Book (two volume edited book: proposal submitted summer 2018 to Routledge)
- Complete literature game on interest (fall 2018)
- Funded by the Office of Naval Research (ONR), Army Research Office, Navy Education and Training Command

Navy Life Game Setting

- A generic [partial] shipboard environment, inside of which, relevant assets, characters, and player capabilities can be outfitted
  - Customizable to different ratings
  - Designed for representing “interesting and challenging slices” of day-in-the-life experiences, rather than showcasing exhaustive, complete suite of tasks & resources
  - Designed for flexibility re: types of measures – e.g. time, interest, engagement, task success, etc.
  - Designed to support data analyses for model development and assessment

Scene:
JOIN Relationship Mapping

		Community									Work Activity																														Work Style									
	Rating	Aviation	Construction	Health Care	Intelligence	Support	Special Warfare/Operations	Submarine	Surface	COUNT	Analyze Communications	Analyze Data	Analyze Documents	Direct Aircraft	Direct Emergency Response	Maintain Documents	Maintain Electrical Equipment	Maintain Electronic Equipment	Maintain Facilities	Maintain Mechanical Equipment	Maintain Security	Maintain Supplies	Maintain Weapons	Make Communications	Make Documents	Make Facilities	Make Mechanical Equipment	Operate Electrical Equipment	Operate Electronic Equipment	Operate Facilities	Operate Mechanical Equipment	Operate Weapons	Respond to Emergencies	Serve Customers	Train People	COUNT		Outdoor	Indoor	Industrial	Office	Physical	Mental	Work independently	Work with team	COUNT	
DC	DC								1	1					1					1													1		1	5		1	1	1	1	1		1	1	7	
FC	FC								1	1					1		1	1													1			1	5			1			1	1		1	4	

SCENE 1 (DC Pipe-Patching):

Scene Construction – JOIN Components

COMMUNITY:

- Surface

WORK ACTIVITY:

- Maintain mechanical equipment
- Operate mechanical equipment

WORK STYLE:

- Indoor
- Work with Team
- Physical

Scene Construction – Game Features

SETTING:

- Ship passageway

REQUIRED TASK:

- Install jubilee patch on leaking pipe

OPTIONAL TASKS:

- Read pipe schematic to locate and close upstream valve
- De-water space using pump
- Compare water pressure upstream and downstream from repair after patch is in place

GAME ASSETS:

- Pipe-patching team (2 NPCs)
- Pipe-patching equipment (jubilee patch, wrench, bag, drop-cloth, pump)
- Leaking pipe (pipe, upstream & downstream valves, pressure gauges)
- Manual with pipe system schematic

Next Steps

- Pre-prototype Navy Life Damage Control game to include metrics and analytics of state-worry interest and self-efficacy
- Validity studies for self-efficacy, test anxiety and interest
  - What is the relationship between game based metrics and learning analytics with JOIN (The gold standard)
  - Does the game increase interest and self-efficacy while reducing test anxiety

Use of Games to Increase Interest

- The 4-phase development model (Hidi & Renninger, 2006)describes how to use game situational interest to develop long-term personal interest to trigger long-term interest in careers in the navy
- Phase 1 or *triggered situational interest* describes the “psychological state of interest that results from short-term changes in affective and cognitive processing”
- Phase 2 or *maintained situational interest* is maintained triggered situational interest. Maintained situational interest is also externally supported such as with the use of instructional conditions
- Phase 3 or *emerging individual interest,* which is the beginning of an enduring predisposition to wanting to repeatedly engage in that task is a transition from externally-supported interest to self-generated interest
- Phase 4 or *well developed individual interest* and is described as more trait-like (enduring and person-generated)

Thank you
Harry O'Neil
honeil@usc.edu

Funded by Defense Advance Research Project Agency (DARPA), Office of Naval Research (ONR), Office of Education (OE), Navy Education and Training Command (NETC)

BACK-UP SLIDES

Affective Measures

- Select vs. Develop
- What is process for selection of construct measures
  - Reliability and validity information
  - Many good choices for trait measures, few choices for state measures (particularly for young children)
- What is the process for development of state affective measures?
  - Adapt existing measures (e.g. for adults)
  - For new measures usability study for ratings via Amazon Turk with adults
    - If adults cannot do the task then kids can not. However, if adults can do the task, kids may not be able to do the tasks
  - Usability study via face to face with kids
  - Revision of items

Interest Description of Ratings (JOIN DNA)

- Damage Controlmen (DC)
  - Navy communities area: surface ships
  - Work activities: direct emergency response, maintain mechanical equipment, operate mechanical equipment, respond to emergencies, train people
  - Work style/work environments: outdoor, indoor, individual, office, physical, work independently
- Fire Controlmen (FC)
  - Navy communities area: surface ships
  - Work activity: maintain electrical equipment, maintain weapons, operate electrical equipment, operate weaponry
  - Work style/environments: indoor, mental work with team

Navy Life Game Context

- Problem
  - The quality and efficiency of recruitment, selection, and retention in the Navy needs to match the expectations and interests of potential recruits for life in the Navy
  - If mismatched, it may result in attrition in basic training and A-Schools. It also means that recruits might not be assigned after basic training to where they can be most productive (e.g., A-Schools and C-Schools)
- Solution
  - Provide potential recruits with a free, motivating, interactive game platform through which they can experience the day-in-the-life of a sailor
  - Special modules could provide recruits exposure to the various jobs, duties and training that exist across different enlisted Navy ratings
  - Provide “real time assessment” to determine the best occupational fit for incoming recruits

Evaluation of the Game

- Evaluation will include:
  - Formative and summative issues
  - Inferences of player knowledge skills, abilities (KSAs), to include interests and self-efficacy
  - Alignment of player KSA and interests to Navy ratings
  - Predicted success of player across various ratings to include predicted attrition and cost avoidance
  - Measuring sailor quality (e.g., ASVAB)
  - Measuring affective attributes

Bibliography

- Chung, G. K. W. K. (2015). Guidelines for the design, implementation, and analysis of game telemetry (pp. 59–79). In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*. New York: Springer.
- Chung, G. K. W. K., & Parks, C. (2015). *Feature analysis validity report* (Deliverable to PBS KIDS). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Madni, A., Chung, G. K. W. K., Baker, E. L., & Griffin, N. C. (2016). Using crowdsourcing as a formative evaluation technique for game icons (pp. 83–98). In H. F. O'Neil, E. L. Baker, & R. S. Perez (Eds.), *Using games and simulations for teaching and assessment: Key issues*. New York, NY: Routledge.
- Marsh, H. W., Hau, K., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psycholog's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. International Journal of Testing, 6(4), 311-360. Lawrence Erlbaum Associates, Inc.
- Paolacci, G., & Chandler, J. (2014). Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184-188.
- Rueda, R., O'Neil, H. F., & Son, E. (2016). The role of motivation, affect, and engagement in simulation/game environments: A proposed model. In H. F. O'Neil, E. L. Baker, & R. S. Perez (Eds.), *Using games and simulations for teaching and assessment: Key issues* (pp. 206-229). New York, NY: Routledge.
- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using mechanical turk to study clinical populations. *Clinical Psychological Science*, 1-8.
- Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2014). Game-based learning. In J.M. Spector, D.M. Merrill, J. Elen, &M.J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 127–222). New York, NY: Springer Science+Business Media.
- Tomas, J., Hesser, H., & Träff, U. (2014). Contrasting two models of academic self-efficacy—Domain-specific versus cross-domain—In children receiving and not receiving special instruction in mathematics. *Scandinavian Journal of Psychology, 55*(5), 440-447.
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van er Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249-265.

Type of Quantitative Research Evidence

- Meta-analysis
  - Quantitative review of the literature
    - See Cohen (1992) and other Cohen references
  - Uses effect sizes
    - A measure of the strength of an effect
    - Description statistic
    - Standardized differences between treatment (inquiry) and control group means divided by the standard deviation
    - Assumes control group is the 50 percentile based on normal distribution
  - A small effect size of .20 would indicate a percentile gain of 8 percentile points or the 58th percentile
  - A medium effect size of .50 would indicate a percentile gain of 19 points or the 69th percentile
  - A large effect size of .80 would indicate a 29 percentile gain or the 80th percentile

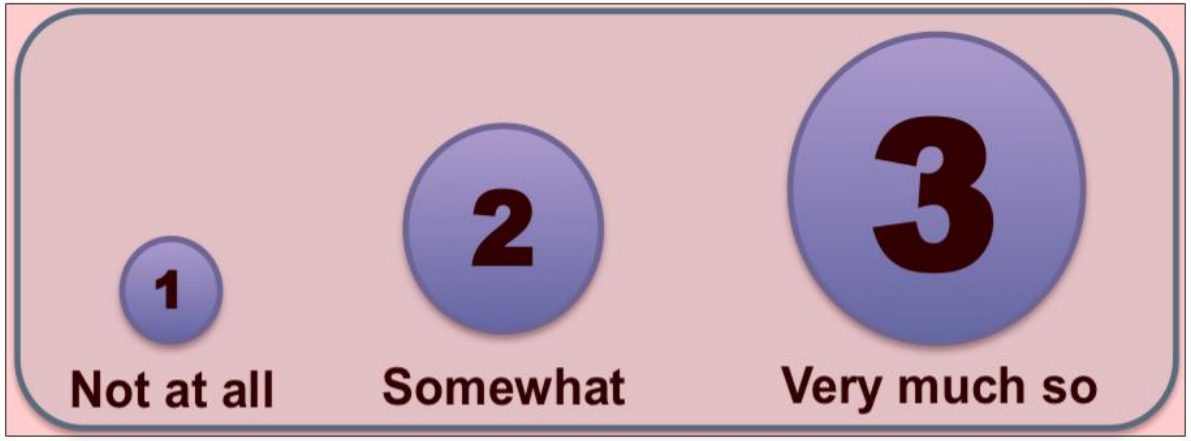
Meta-analysis Effect Sizes

	Media Comparisons	Value Added
Clark et al., 2015	0.33	0.34
Wouters et al., 2013	0.29	--
Tobias et al., 2015	0.63 (Cognitive) 0.49 (Affective)	-- --

Amazon Turk Instructions

- We are developing icon/image answer options to be used with young children when they are answering questions about their experience playing an educational physics game. We need your help to determine which options are best. Below are example instructions and a question that the young children might be asked
- “Tell me how you thought or felt about this game”
- “I’m certain I mastered the skills being taught in the game”
- Look at the answer icons/images below and select the one that you think would work best for young students to use when answering questions about the intensity of their feelings, such as the example question provided above

State Intensity Icons *(N = 14)*

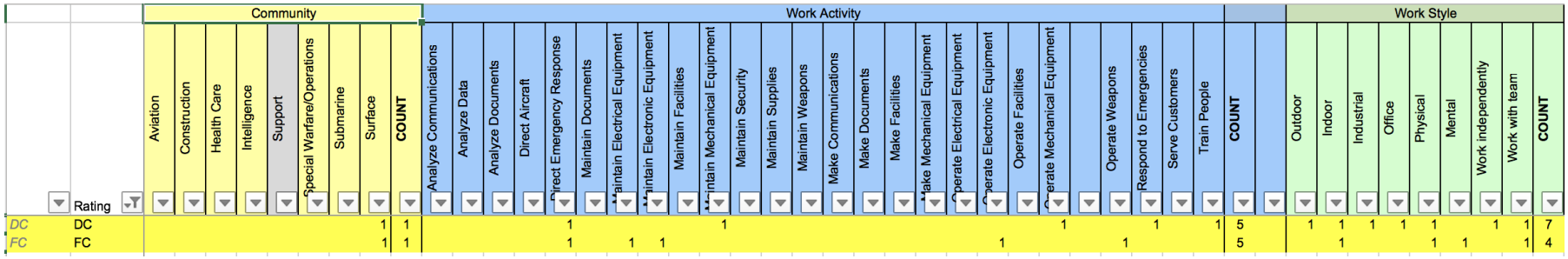


- Intensity circle icon: 10/14 responders
- Intensity bar graph icon: 0/14 responders
- 4/14 nonresponders

Game Research Context: Navy Life

- Navy jobs (ratings)
  - Damage Controlmen
  - Fire Controlmen
- New Process
  - Background info then ASVAB (1ST), ITAB (2ND), JOIN (3RD), Game-based measures (4TH)

JOIN Parameters: Differences between DC and FC Ratings



		Community								Work Activity																												Work Style									
	Rating	Aviation	Construction	Health Care	Intelligence	Support	Special Warfare/Operations	Submarine	Surface	COUNT	Analyze Communications	Analyze Data	Analyze Documents	Direct Aircraft	Direct Emergency Response	Maintain Documents	Maintain Electrical Equipment	Maintain Electronic Equipment	Maintain Facilities	Maintain Mechanical Equipment	Maintain Security	Maintain Supplies	Maintain Weapons	Make Communications	Make Documents	Make Facilities	Make Mechanical Equipment	Operate Electrical Equipment	Operate Electronic Equipment	Operate Facilities	Operate Mechanical Equipment	Operate Weapons	Respond to Emergencies	Serve Customers	Train People	COUNT		Outdoor	Indoor	Industrial	Office	Physical	Mental	Work independently	Work with team	COUNT
DC	DC								1	1					1						1										1		1		1	5		1	1	1	1	1		1	1	7
FC	FC								1	1					1		1	1												1		1				5			1			1	1		1	4

Rating Distinctions within JOIN

Parameters Unique to DC Rating:

Parameter	Type
Maintain Mechanical Equipment	Work Activity
Operate Mechanical Equipment	Work Activity
Respond to Emergencies	Work Activity
Train People	Work Activity
Outdoor	Work Style
Industrial	Work Style
Office	Work Style
Work Independently	Work Style

Parameters Common to Both

Parameter	Type
Surface	Community
Direct Emergency Response	Work Activity
Indoor	Work Style
Physical	Work Style
Work with Team	Work Style

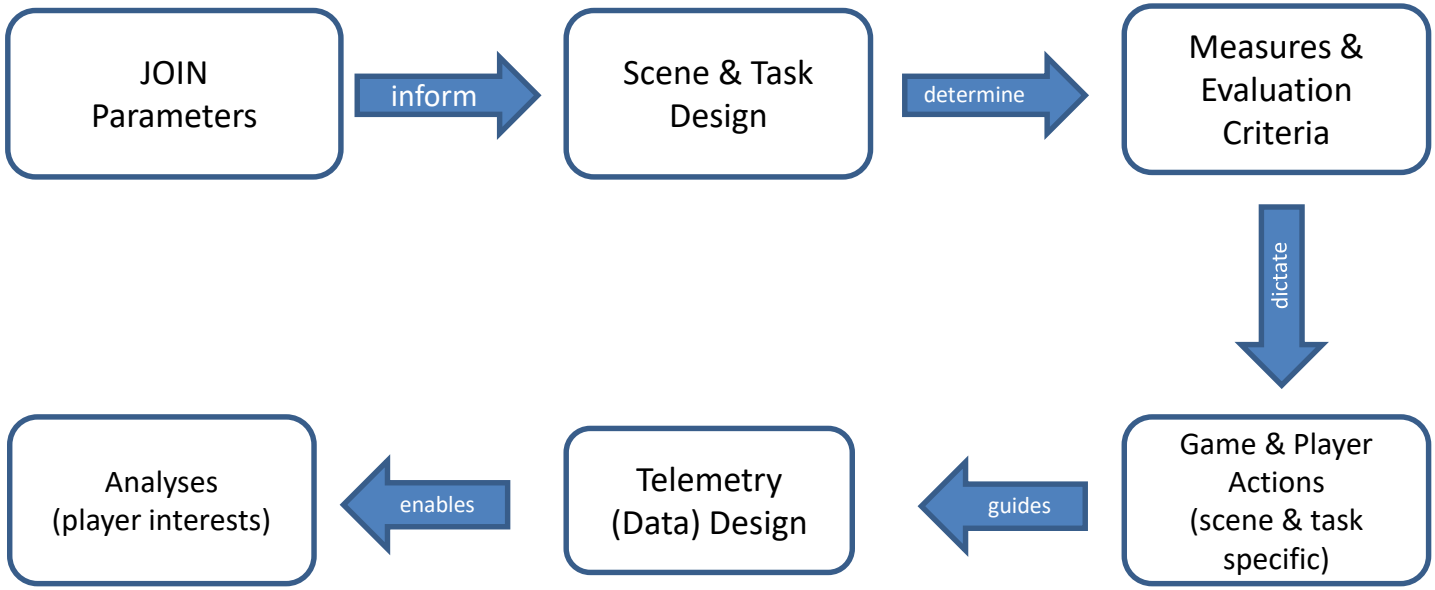
Parameters Unique to FC Rating:

Parameter	Type
Maintain Electrical Equipment	Work Activity
Maintain Electronic Equipment	Work Activity
Operate Electronic Equipment	Work Activity
Operate Weapons	Work Activity
Mental	Work Style

Measure of Interest – JOIN

- The Jobs and Occupational Interest in the Navy (JOIN) is a computer-administered inventory of a recruit's vocational interest
  - Designed to use pictorial representations of the job descriptions to alleviate the issue of recruits not knowing what the job entails
  - Each item on JOIN was accompanied by four pictures representing that item with behavioral descriptions of the tasks performed
  - Recruits are asked to indicate on a 5-point Likert scale their interest level (very interested, neutral, and not interested)
  - JOIN is organized by Navy community areas (aviation, construction, health case, intelligence, submarine, surface, special programs and support), work styles (mental, physical, work independently, work with a team), work environments (indoor, outdoor, industrial, and office), and work activity items (these are more specific such as analyze data, direct emergency response, etc.)
  - Evidence of reliability and validity of JOIN suggest acceptable alpha coefficients (alpha for the whole scale was .91 and ranged from .83 to .95 for each work activity)

Game Design Work-Flow



Interest in Navy Context of Selection and Classification