Expert Similarity Index (ESI):
A Serious Games Analytics Performance Index

ORGANIZATIONAL SCIENCE TRACK

Christian "Sebastian" LOH, Ph.D.
Southern Illinois University

EFFECTS OF ANALYTICS ON SERIOUS GAMES

Analytics create new opportunities (market) for industries/research

- Analytics is contextual → Asking the right questions for the field (no need to copy others)

The effects of analytics on games

- MMO game analytics → monetization (advanced data science)
- Commercial games: ‘No’ analytics openly, Leader’s Board (highest kill, fastest time)
  - Real analytics is safe-guarded for internal evaluation (beta testing) to improve game

Serious Games (Education vs Organizational Science)

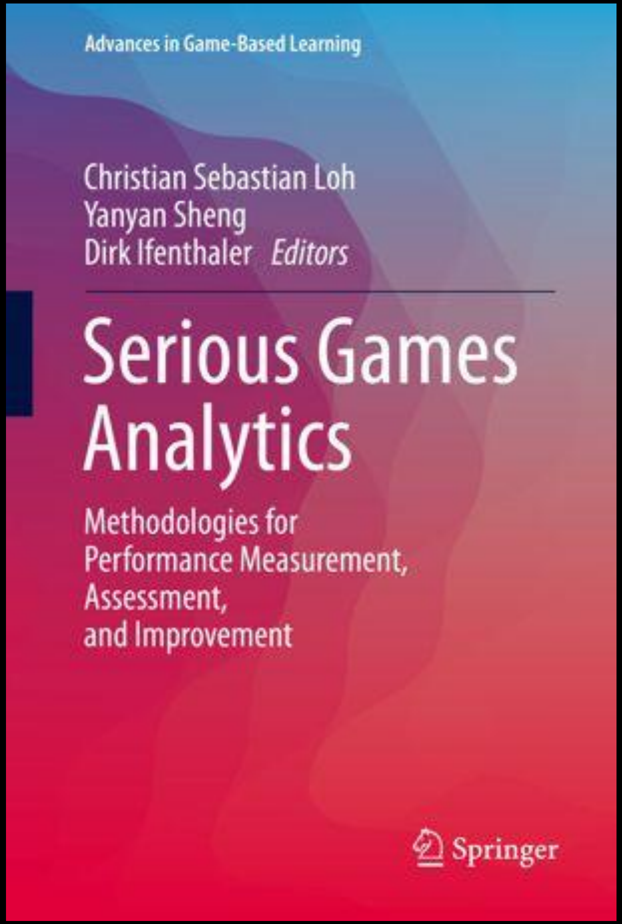
- NSF-workshop offers two tracks: Education vs Organizational Science
- (1) DGBL (Education use): “No assessment, fun learning” (argument is moot)
- (2) Edutainment Reboot? – Assessment looking more like learning analytics (similar to LMS)
- Serious games for Organizational learning (should take different approach from DGBL)

SERIOUS GAMES ANALYTICS RESEARCH

Serious Games Analytics: Methodologies for Performance Measurement, Assessment and Improvement – [Eds] Loh, Sheng, & Ifenthaler (2015)

- Loh, et al (2007),
- Loh (2012), (2013)
- Loh & Sheng (2013), (2014), (2015)
- Loh, Sheng & Li (2015)
- Loh & Ekstrand (2015)
- Loh, Li & Sheng (2016)
- Loh & Li, (2010), (2015), (2016)

- Papers available on http://www.csloh.com/research and Research Gate



WHAT I HAVE LEARNED AFTER 10 YEARS

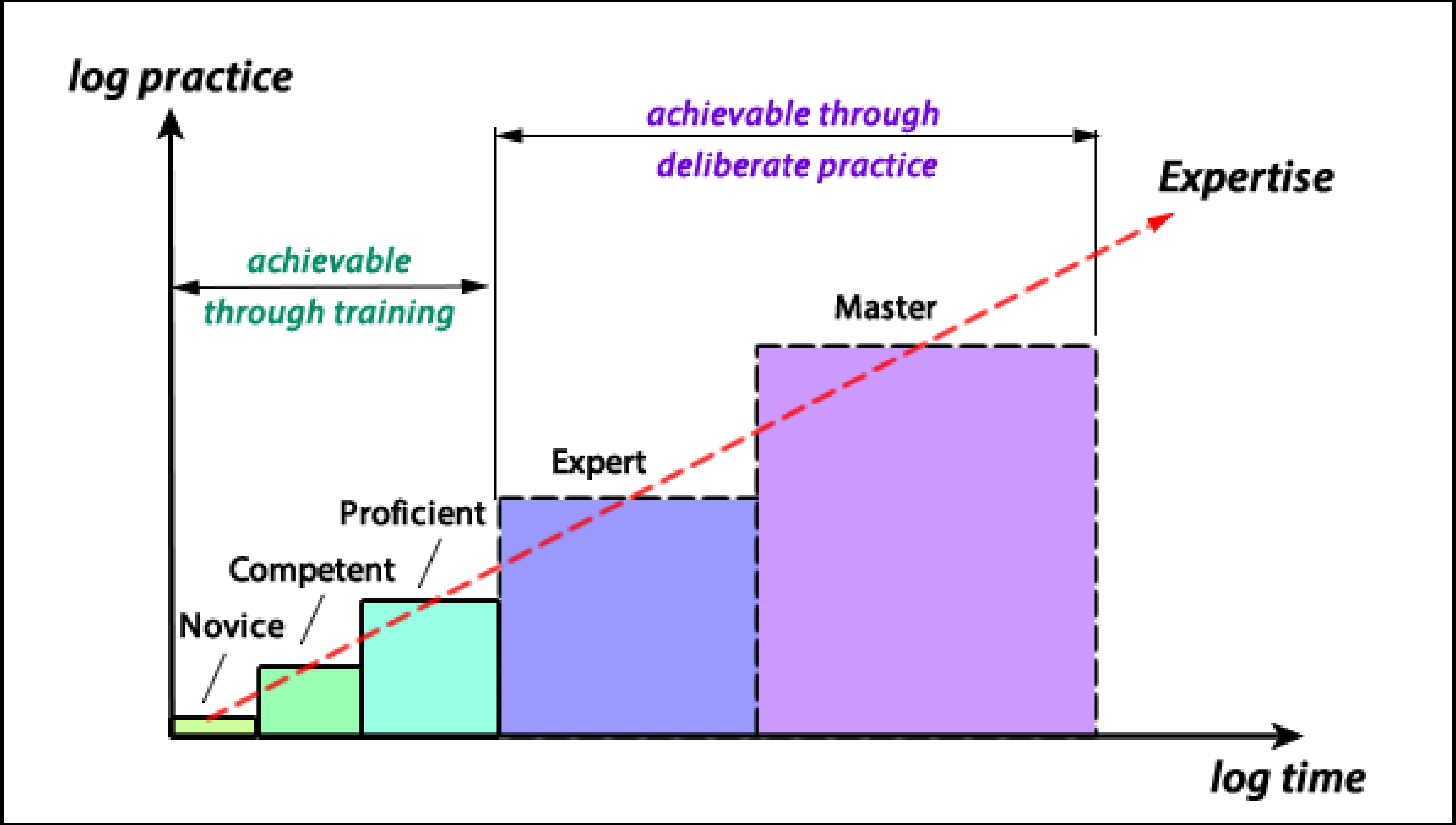
(of designing our own serious games for research…)

Not in order of 'discovery' learning:

Wasted time: Once-off (serious) games – most common – is not very useful

- Perusal learning, did not encourage practice/training, NOT ENOUGH DATA for analytics

1. Use repeat play (multi-rounds) in games to increase metacognition
   - Deliberate Practice – metacognitive activities (practice to proficiency and mastery)
2. Serious games training (with repetition) apparently follows 'Learning Curve'
3. Levels of Skill Acquisition
   - Novice – Competent – Proficient
     - Trainable (short term) – practice to mastery (with repetition)
   - Expert – Master
     - Not directly trainable, require Deliberate Practice over time (~10 years)
4. Viable to compare Novice's (course of) actions to Expert's as performance assessment!



Skill Acquisition + learning Curve + Deliberate Practice (with Instructional Technology)

PRIOR SUPPORTING RESEARCH

For your reference

PRIOR (SUPPORTING) RESEARCH

1. Learning Curve (Thurstone, 1919)
   - Serious games (learning) performance follows a Learning Curve
   - Professional gamers: 8-10 hrs of practice everyday to maintain performance
   - Repetitive serious play (procedural) is NOT the same as rote learning (declarative) in edutainment

2. Deliberate Practice Develops Expertise (Ericsson, 2006)
   - DP: Effective improvement of performance that entails sequential, mindful repetitions of a training task, along with immediate feedback, such that expert performance is acquired gradually
   - ~10,000 hours (10 years) of DP to achieve “Mastery” (e.g., Chess Master, Music Maestro)
   - Repetitive play leads to valued behaviors (skills in doing certain tasks)

3. 5-Level of Skill Acquisition (Dreyfus & Dreyfus, 1980)
   - Novice → Competent → Proficient → Expert → Master
   - Novice to Proficient is trainable (short term), Expert to master required DP (long term)
   - Minimum 100 hours to acquire ‘proficient’ cognitive skill (Anderson, 1982)
   - Repetitive play shortens the time taken to achieve proficiency in valued behaviors

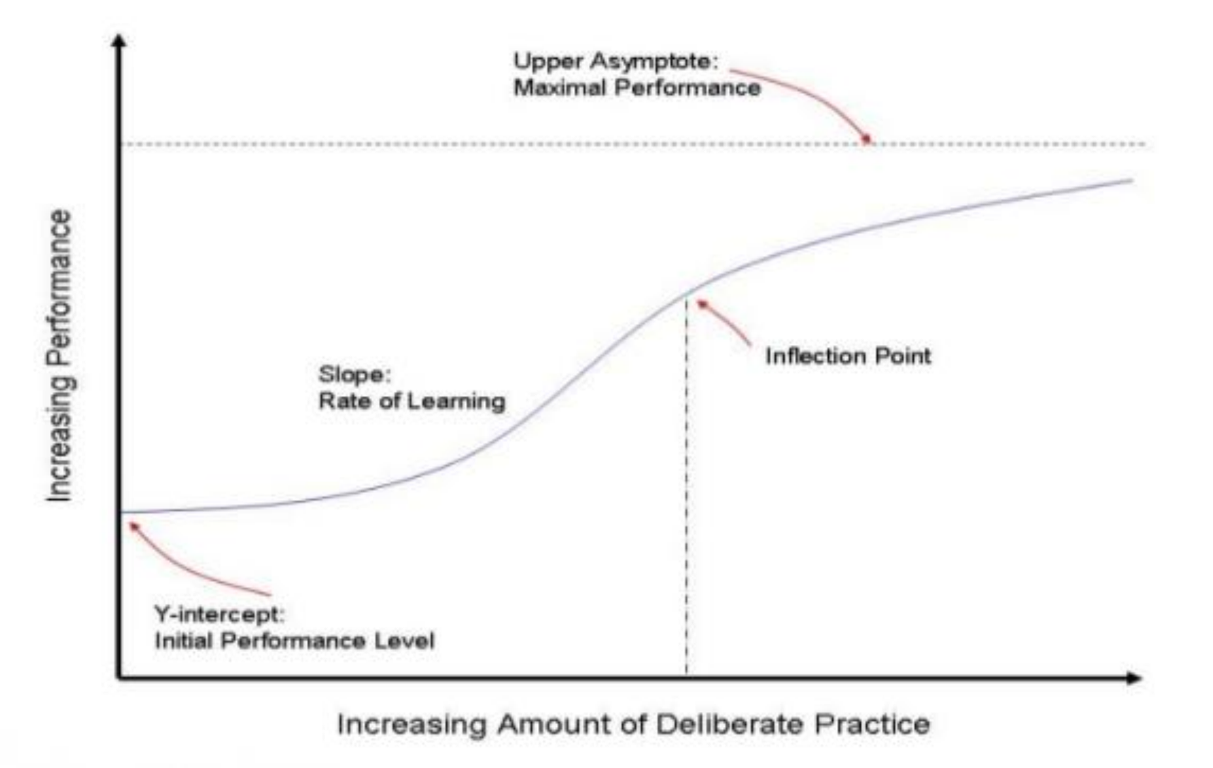
1. LEARNING CURVE

When performance is plotted against (continuous) time spent on learning (DP)

- LC equation first reported by Thurstone (1919)
- More useful as a measurement (descriptive) than for performance improvement (prescriptive)
- LC is a sigmoidal curve
- But Log (practice) - Log (Time) will produce a straight line

Research: measure 'how much training is necessary' in medical education

- Predict 'how much practice is required to achieve competency' in medical surgery
- Pusic, Pecaric, Boutis (2011)



2A. DELIBERATE PRACTICE (DP)

Useful instructional strategy for the development of expertise (Ericsson, 2006)

- Effective improvement of performance that entails sequential, mindful repetitions of a training task, along with immediate feedback, such that expert performance is acquired gradually
- Approximately 10,000 hrs (~10 years) of DP to achieve "Mastery" (Chess Master, Music Maestro)
- Min. 100 hrs needed for cognition skill acquisition and practice to reach proficiency (Anderson, 1982)

Serious games should be designed to include 'repetition'

- Design repetitive game events to afford more metacognition and procedural learning
- Metacognitive activities: Self-correction, reflection, decision making, strategizing
- Promote decision-making and practice-to-mastery, build memory, learn from mistakes, improve strategy
  - Can you do this series of events better (in less time, with better accuracy, etc)
- Once-off game is 'less useful' for training (only peruse the process, NO practice)

2B. MEASURING NOVICE-EXPERT BEHAVIORS

Novice-expert behavioral differences are well-established (Ericsson, 2005)

- (Sufficient) training will help novices bridge performance gap to become more like experts
- Serious games (as a form of training) can help novices become more like experts
- Select/identify suitable game activities (to train novices → more like experts)

How to measure people's learning process in serious games for assessment?

- Track the learning process to verify if serious games maximize value-adding behaviors
- Evidence of the learning ought to be measurable, replicable, or demonstrable
- Desirable – a metrics (standardized, generalizable) useful for identifying valued behaviors
- Expert(s)' performance is the desired outcome (value-added behaviors) to be emulated
- Compare novices' performance against that of the expert(s) as feedback for metacognition

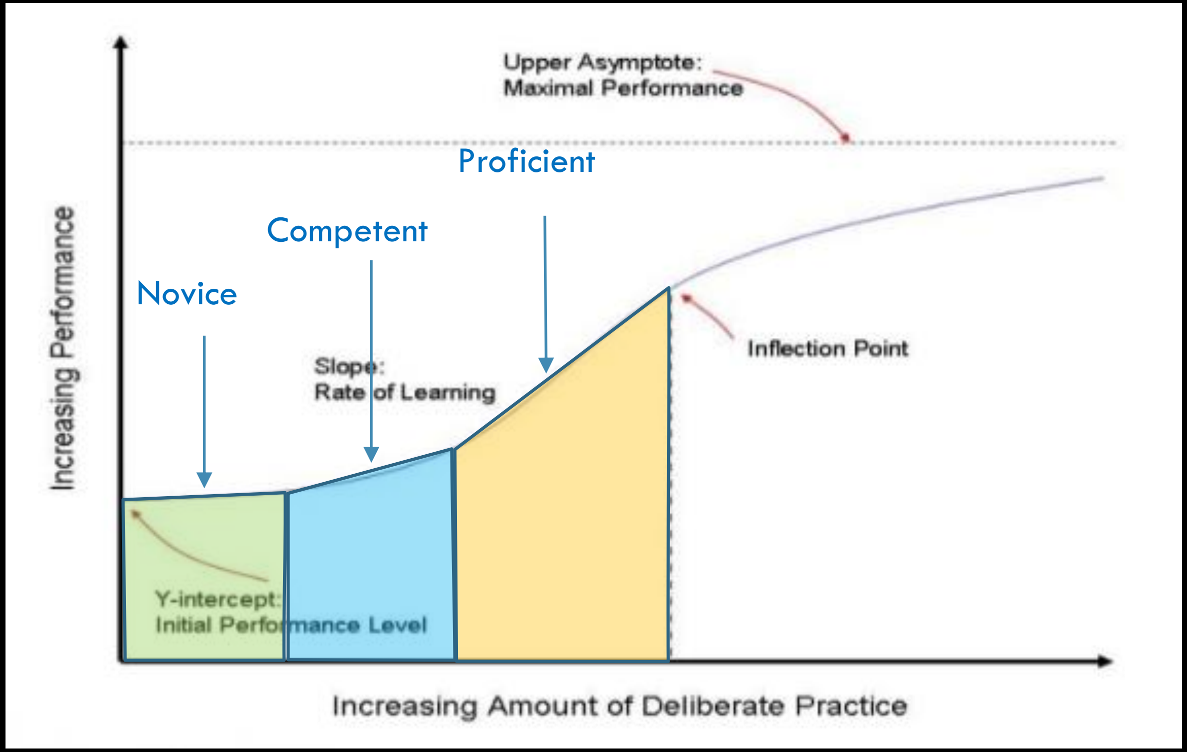
3A. NOVICE TO PROFICIENT (SKILL ACQUISITION)

Loh, Li, Sheng (2016)

1. Novice – initial inertia, may give up
2. Competent – noticeable improvement
3. Proficient – biggest improvement

First 3 stages achievable through Serious Games (short duration, suitable for research)

Proficient can be good teacher to 1, 2.



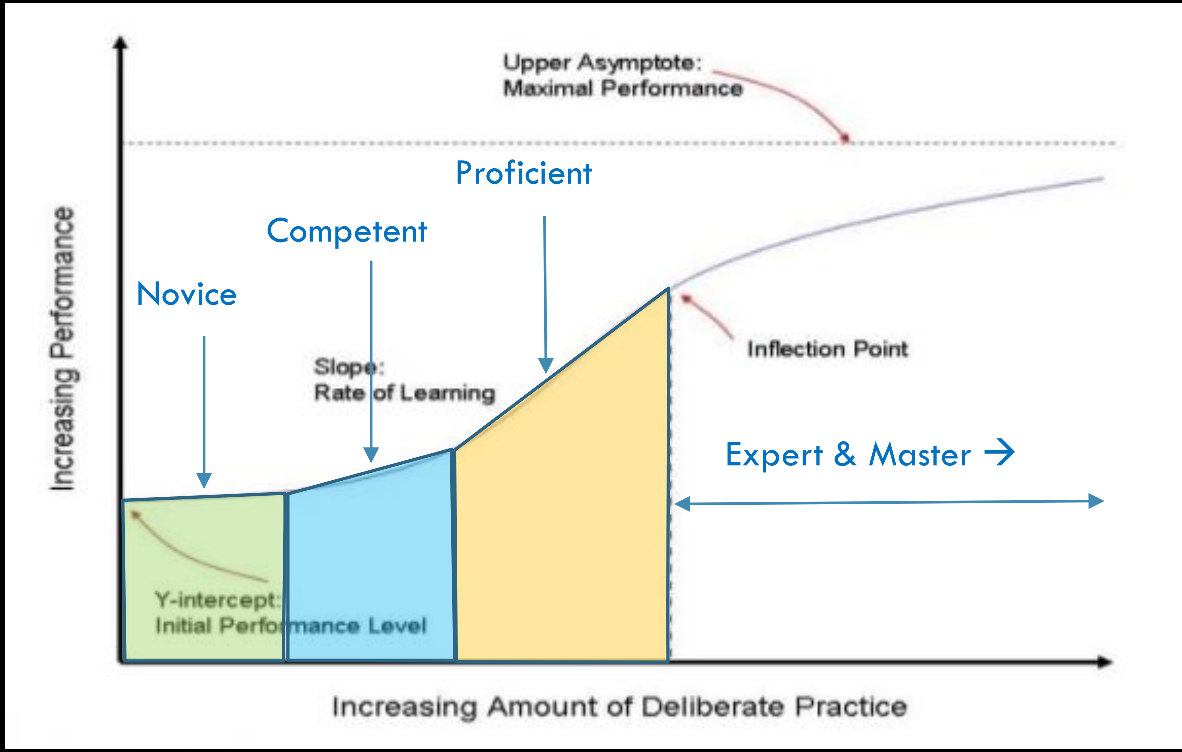
3B. EXPERT AND MASTER (SKILL ACQUISITION)

Loh, Li, Sheng (2016)

1. Novice – initial inertia, may give up
2. Competent – noticeable improvement
3. Proficient – biggest improvement
4. Expert – require DP (and time)
5. Master – require DP (~10,000 hrs)

Proficient & Expert can be good teachers
Masters are not good teachers

Last two stages achievable only through DP, can takes years (not suitable for SGA research)



BACK TO PRESENTATION…

GAMING, PLAYING, LEARNING BY DOING

Serious games for organizational performance fits "Learning by Doing"

- Learning by Doing (Procedural Learning) → 'practice' leading to mastery
- How to perform tasks, What is the 'sequence of events' to be carried out in order
  - E.g.: Medical (surgery/anesthesia), Computing and Engineering (cybersecurity, programming, manufacturing), Flight Operation, Sport Science/Martial Arts, any performance where "procedures" is key

Games comprise of "a series of engaging purposeful actions" (play)

- DGBL: more declarative (learning about 'things') than procedural
- Organization focus more on skill acquisition: more procedural knowledge
  - A series of events/activities aimed to promote (semi-)permanent change of behaviors
  - Activities for metacognition: decision-making, strategizing, (self-)reflection, (self-)correction…

VALUED BEHAVIORS FOR ORGANIZATION

Actions can: (1) add value, (2) add no value/bring risks or losses

- Actions (what people do) can be: learned (encouraged) AND unlearned (discouraged)
- Experts' actions are valued, and to be emulated by novices
- Novices in organization should act more like experts over time with training
    - Practice to achieve mastery (really, proficiency)

Serious games should MAXIMIZE value-adding behaviors in the organizational learning process

- Encourage valued actions (learned actions if not exist, strengthened if weak)
- Put value tags on people's actions (in the training, just like workplace)
- Performance to be based on people's actions (process-based, fits Learning by Doing)
- Verify that novices learned 'valued actions' similar to that of the experts

SERIOUS GAMES SHOULD…

Employ "repeated sessions" (to encourage practice and metacognition)!

- Game activities should not be rote learning (boring), but metacognition
- Metacognition → Action → Behavior (Muscle memory IS cognitive memory!)
- Design activities for value-added behaviors, opportunities for metacognition

ASK: Can you perform actions/tasks up to a predetermined level?

- Very different methods to determine (1) 80% of assessment (score) vs. (2) 80% like the expert
- (1) Pretest/Posttest method – most prevalently reported method (easiest?) in SGA research (Bellotti, et al, 2013) – more suited for asking declarative knowledge
- (2) Better: (more difficult) Determine if someone is 80% like an expert

Design repetition (procedural) into serious games (NOT Rote Learning – declarative)

- Repetition with high metacognitive activities to 'train' novices to think and behave more like experts
- $$ 'Repetition' session data needed for Predictive Analytics – forecasting!
- Predictive Analytics: shorten training time, know when to stop training → cost saving!

MEASURING NOVICE-EXPERT BEHAVIORS

(Sufficient) Training will help novices learn to become more like 'experts'

- Serious games – a form of training (practice will increase performance → like experts)
- Identify/select suitable metacognitive game activities (to train novices → more like experts)

Measure in-process learning (not just outcome) for serious games assessment

- KPI (outcome based indicators, summative) vs PPI (Process Performance Indicators, formative)
- Evidence of the learning ought to be measurable, replicable, or demonstrable
- Track the in-game learning process to verify if novices 'pick up' value-adding behaviors
  - Telemetry – in-process data (MySQL, R, data mining, machine learning)
- Desirable – an index/metrics (standardized, generalizable) to measure valued behaviors
- Use expert performance as baseline for PPI comparison (rather than pretest/posttest, KPI)

CAPTURING THE PROCESS DATA

Current methods to capture process data

- “After Action Report”, Feedback, Screen Capture, Usability, Human Performance, User Experience, Body/Action Cam, etc.
- Largely qualitative – requiring expert judgement, training for consistency, eye-balling

Disadvantage of Outcome-based (only) assessment

- Do not consider cheating (can ‘achieve’ some outcome, but increase risk)
- Missed out on new/improved process (can alter outcome)
  - Examples: Flight accidents, robotic surgery errors

Quantitative approach (data science)

- Process lots of data (scalable for use by organizational training)
- Objective (data-driven) rather than subjective interpretation of performance
- ESI suitable for both *post hoc (KPI)* and *ad hoc (PPI)* analysis

TELEMETRY

Log files can capture in-game processes, events and action data

- Log files and After Action Reports are not enough – *post hoc* ONLY analysis
- Patterns discoverable through machine learning/data mining/visualization, etc.

Telemetry (live data) → Data Science (best option)

- Remote data sensing/collection with *in situ signal* transmitters and receivers
- Requires server to trace user-generated (interaction) data: real time, live (*ad hoc*) analysis
- Process Performance Indicator (PPI): in-process analytics
- More versatile, direct incorporation into machine learning workflow → MySQL, R

Capture both outcome AND in-process data!

EXPERT SIMILARITY INDEX (ESI)

A Serious-Games Performance Analytics

EXPERT'S ACTIONS (VS NOVICES)

Assumption: Experts' actions are better (the best/ideal route)

- NOTE: (True) Experts (>10 years of DP) – difficult to find
- 'Organizational' (identified) Experts and Likely Experts (also high proficient and game designers)

Steps: *Serious Game: learning environment containing suitable training scenarios (repetition)*

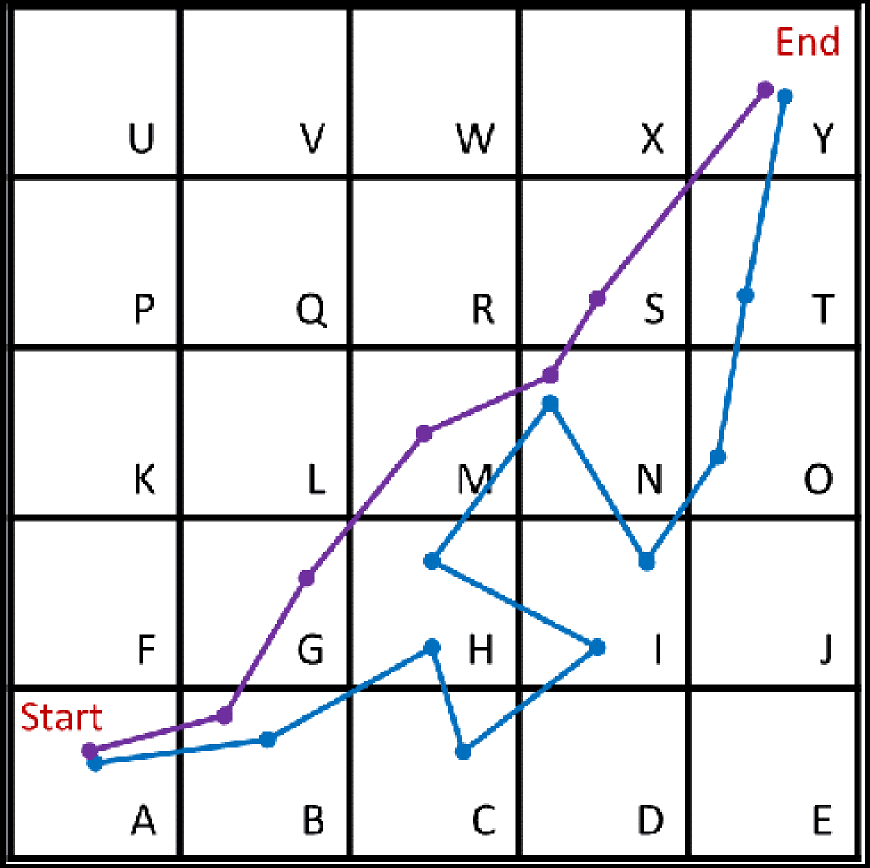
- Identified 'expert(s)' to experience training scenarios (repetition) to establish baseline

PHASES OF EXPERT LEARNING (why repetition is necessary):

1. Exploration Phase: trying out, familiarization of environment
2. Metacognition Phase: thinking/strategizing best ways to problem-solve
3. Self-correction Phase: testing out various strategies to identify best solution

- Expert's 'Course of Actions' (COAs) = 'Best'/Ideal Outcome(s)

EXPERT'S "COURSE OF ACTIONS" (COAs) [1]



Expert: ABGMNSY

Novice: ABHCIHNIOTY

!! Course of Actions (COAs) is essential for the calculation of ESI

SIMILARITY INDICES (STRING/TEXT EXAMPLES)

Jaccard Index: $\left|\frac{A \cap B}{A \cup B}\right|$

(1) POPS vs STOPS

PO, OP, PS | ST, TO, OP, PS

- Intersects (2: OP, PS)
- 'Unique' unions (5: PO, OP, PS, ST, TO)

$\text{Jaccard similarity index} = \frac{2}{5} = 0.4$

(2) POPS vs POPS (PO, OP, PS)

$\text{Jaccard similarity index} = \frac{3}{3} = 1$

- Identical, completely similar

Dice Index: $2\frac{|A \cap B|}{|A|+|B|}$

"I am here" | "here I am, not"

- I am, am here | here I, I am, am not
- Intersect (1: 'I am')
- |A|(2: I am, am here)
- |B| (3: here I, I am, am not)

$\text{Dice similarity index} = 2 \text{ x } \frac{1}{5} = 0.4$

'here I am, not' vs 'here I am, not'

$\text{Dice similarity index} = 2 \text{ x } \frac{3}{6} = 1$

EXPERT'S "COURSE OF ACTIONS" (COAs) [2]

'Course of Actions' can include:

1. (Location-based) Coordinates (x, y): (x1, y1, z1), (x2, y2, z2),…
   - Game world vs real world coordinates – can use (Arc)GIS
2. Events (A,B): (x1, y1, A), (x2, y2, B),…
   - Use of *n*-gram to incorporate 'directionality' in Course of Actions
3. Time component (t): (x1, y1, t1), (x2, y2, t2),…
   - Leaderboard concept: less useful from our experience
   - However, consider that with 'sequences of events', it already subsumes the time component

!! Course of Actions (COAs) is essential for the calculation of ESI

"DIRECTIONALITY" AND *N*-GRAMS

Directionality of COAs is established/preserved with *n*-grams

String: ABCDEFG
n-gram = 1 : A, B, C, D, E, F, G [7]
n-gram = 2 : AB, BC, CD, DE, EF, FG [6]
n-gram = 3 : ABC, BCD, CDE, DEF, EFG [5]
n-gram = 4 : ABCD, BCDE, CDEF, DEFG [4]
…

Study on how different *n*-grams improve/affect ESI (Loh, Li, Sheng, 2016)

- Machine Learning: *StringDist* package for R (van der Loo, 2014)

STRING SIMILARITY INDEX

Similarity Index (or, coefficient)

- Many indices for different purposes: e.g., string comparison, text mining, DNA sequencing
- DO NOT MIX – Each has different strengths and weaknesses (Goshtasby, 2012; Cha, 2007)

Expert Similarity Index:

- Standardized metrics to measure procedural learning/knowledge (i.e., PPI)
- Index: Ranges from 0 (completely dissimilar) to 1 (identical)
- Metrics: Accurately and meaningfully compares similarities between two COAs
- Comparison of common similarity indices as ESI (Loh, Li, Sheng, 2016)
  - Dice, Jaccard, Overlap, Cosine, Longest Common Substring

$$Dice(A,B) = 2\frac{|A \cap B|}{|A| + |B|}$$

$$Jac(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

$$OVL(A,B) = \frac{|A \cap B|}{\min(|A|,|B|)}$$

$$Cos(A,B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

$$LCS(A,B) = 1 - \frac{d_{LCS}(A,B)}{d_{max}(A,B)}$$

VARIATIONS OF ESI: MAXIMUM, MINIMUM,…

ESI = COAs comparison between unknown (novice) vs known (best) routes

- (n) Novice(s) to (1) Expert (Loh & Sheng, 2013)
- (n) Novice(s) to (n) Experts (Loh & Sheng, 2014)

We added several variations of ESI

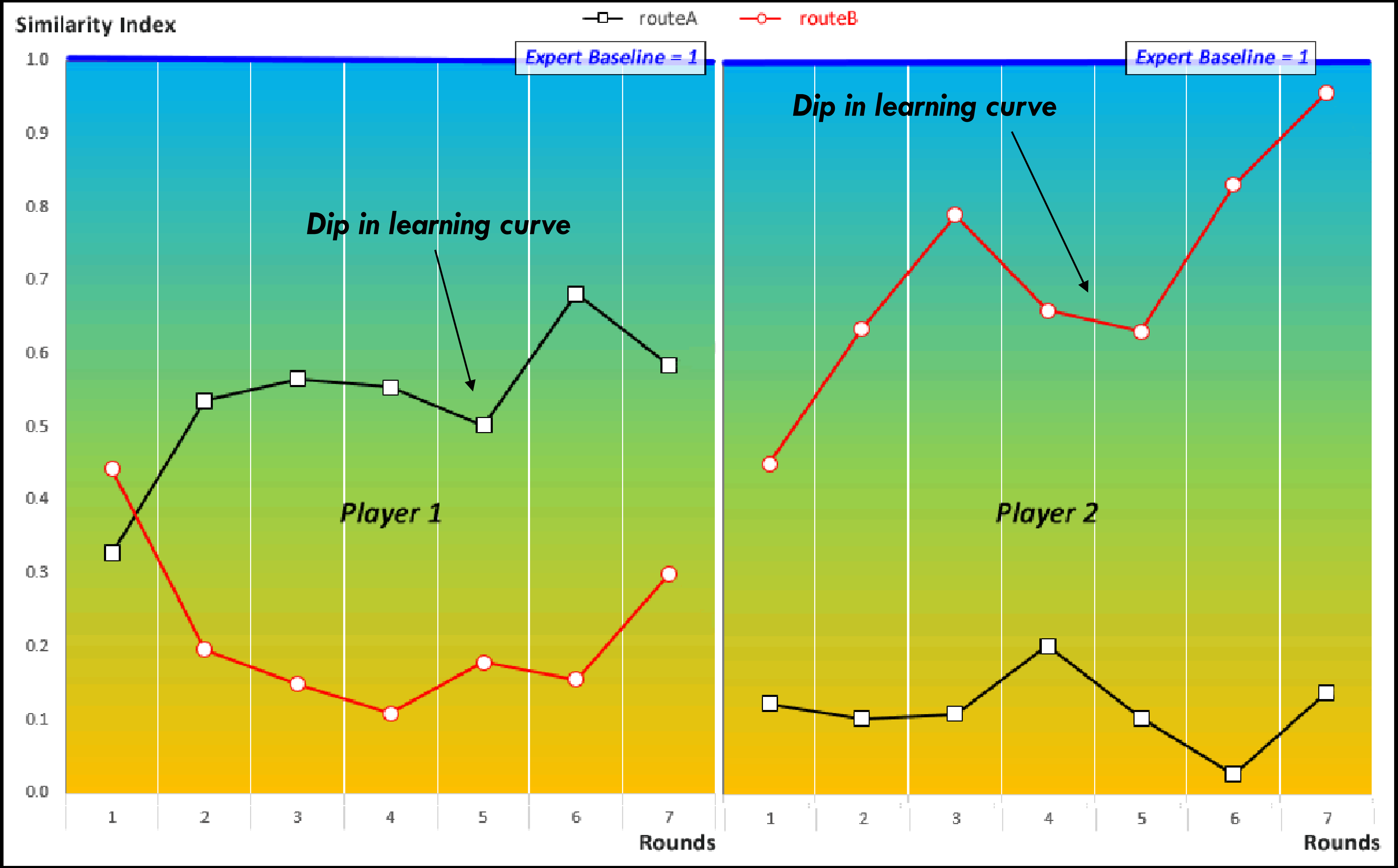
- Presence of multiple experts' – requires Maximum Similarity Index (MSI)
  - See Loh & Sheng (2014) for explanation and computation method for MSI
  - Other possible ESIs: Minimum SI, Average SI, etc. (see Loh, Li, Sheng, 2015)

Improve fit of discriminant analysis

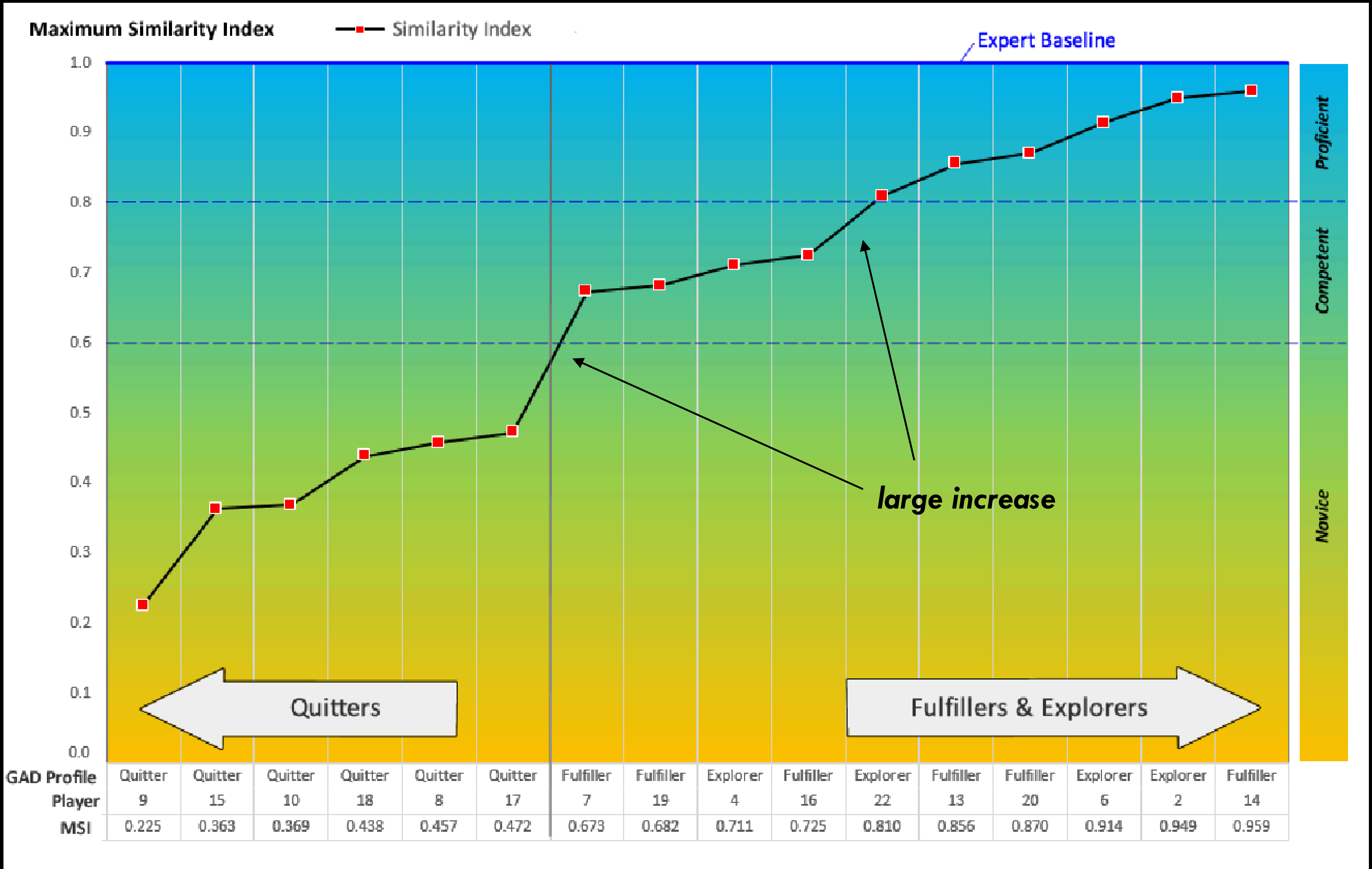
- Using MSI alone (77.4%, $R^2 = .58$), Combination various SIs (93.6%, $R^2 = .91$)

!! Course of Actions (COAs) is essential for the calculation of ESI

ESI VISUALIZED…

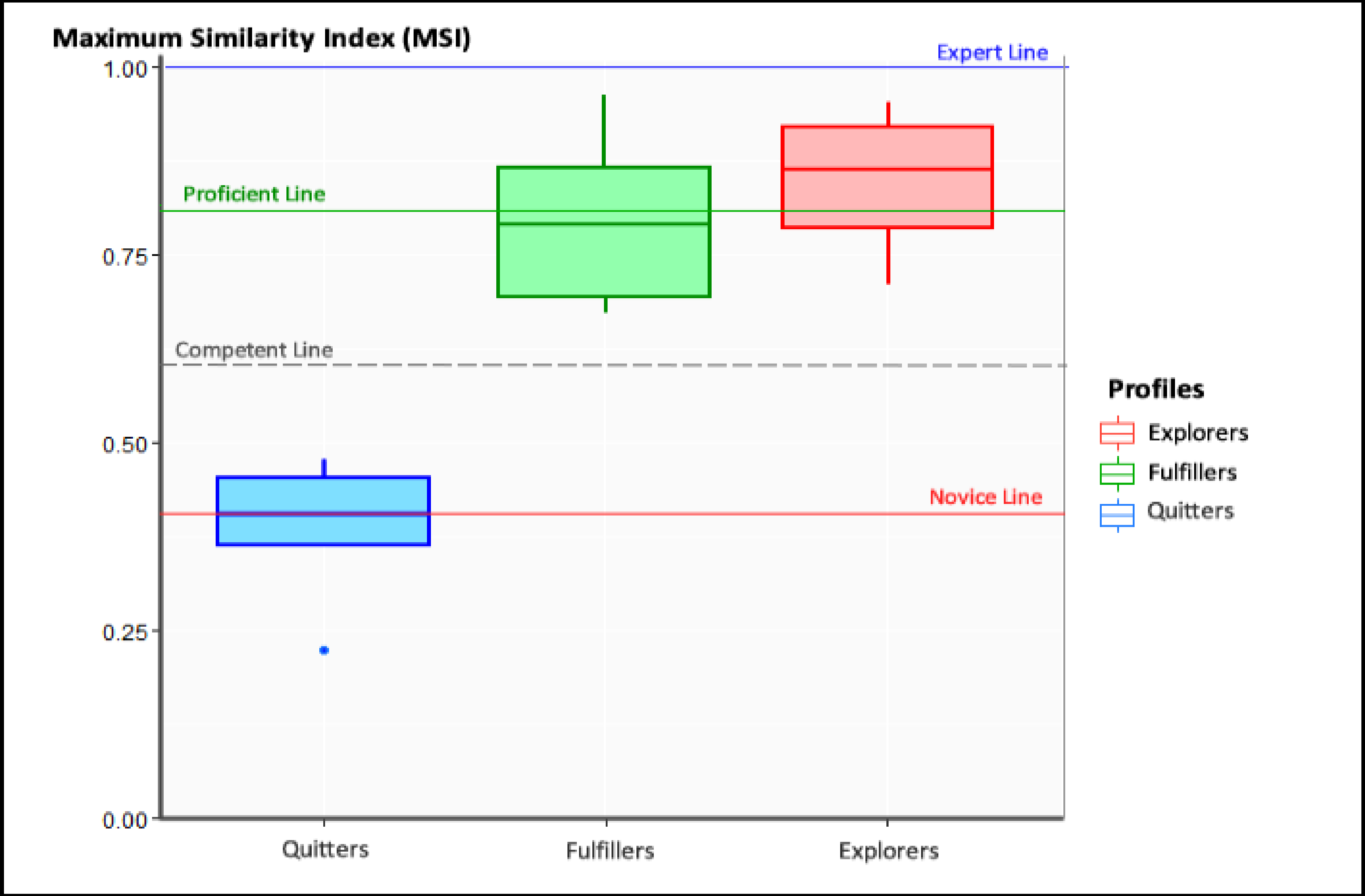


Multiple (2) expert routes (performance increase 'obeys' learning curve)

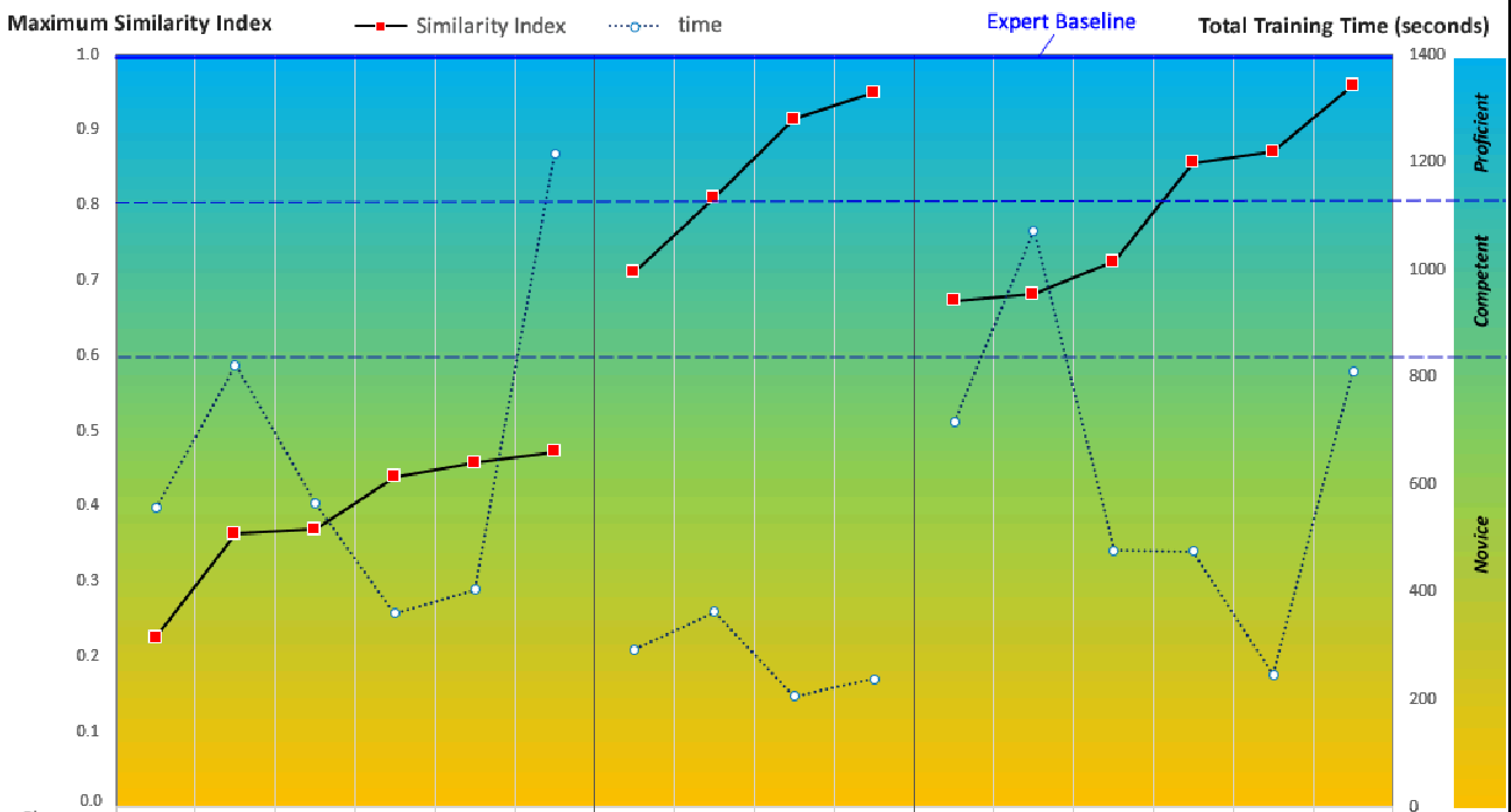


GAD Profile	Quitter	Quitter	Quitter	Quitter	Quitter	Quitter	Fulfiller	Fulfiller	Explorer	Fulfiller	Explorer	Fulfiller	Fulfiller	Explorer	Explorer	Fulfiller
Player	9	15	10	18	8	17	7	19	4	16	22	13	20	6	2	14
MSI	0.225	0.363	0.369	0.438	0.457	0.472	0.673	0.682	0.711	0.725	0.810	0.856	0.870	0.914	0.949	0.959

Maximum ESI: Creating trainee profiler



Maximum ESI: Identifying Novice, Competent and Proficient



Player	9	15	10	18	8	17	4	22	6	2	7	19	16	13	20	14
GAD Profile	Quitter						Explorer				Fulfiller					
Max Sim Ind	0.225	0.363	0.369	0.438	0.457	0.472	0.711	0.810	0.914	0.949	0.673	0.682	0.725	0.856	0.870	0.959
Max time	556	822	564	361	404	1216	293	363	206	238	716	1073	478	476	246	811

Time of Completion (used in Leader Board) not a good indicator of performance

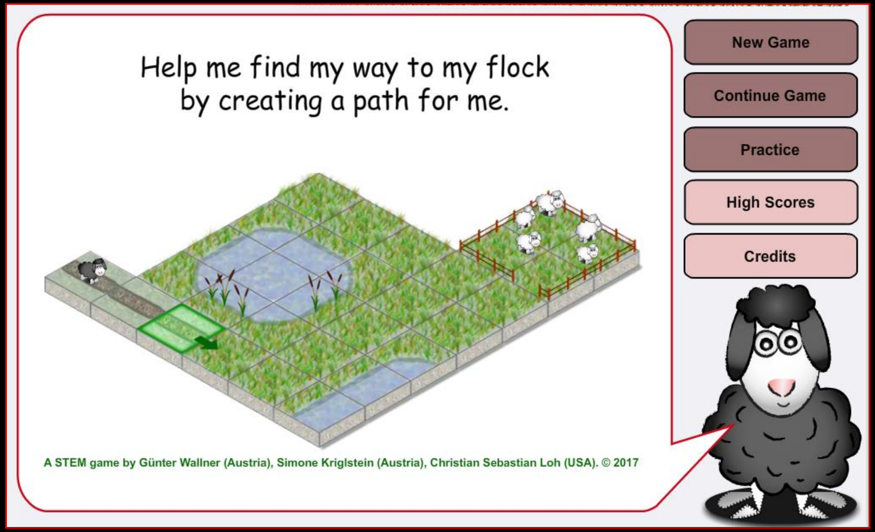
BUGS IDENTIFICATION

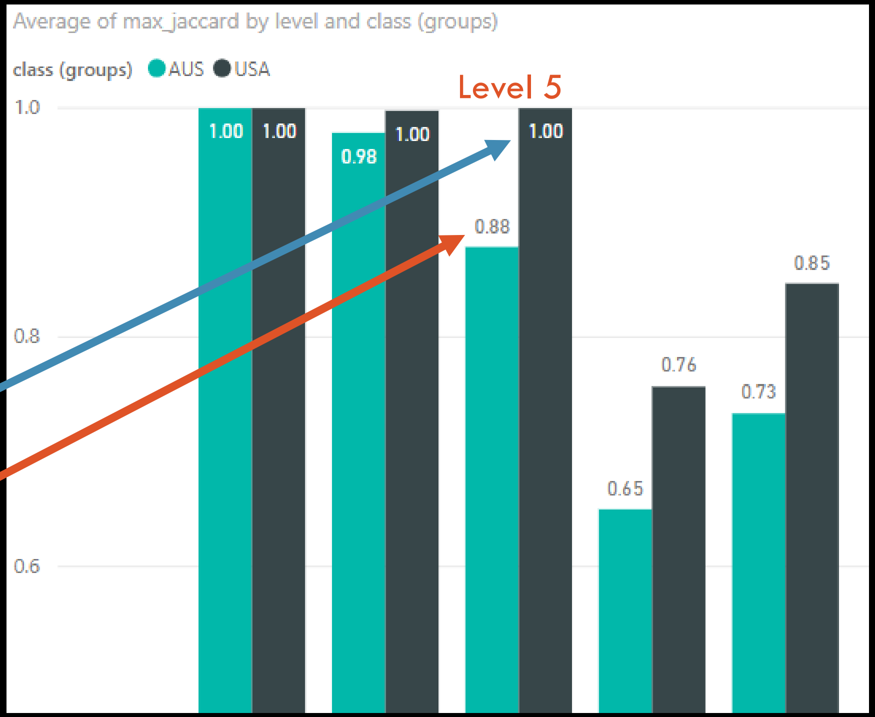
Research: Puzzle-solving geometric game

- Limited resources: given fixed number of tiles to build road
- Players must solve puzzle (complete the road) to move to the next level
- Expectation: Only 1 way to clear the level, everybody should get ESI of 1.0

Level 5 only 1 solution (ESI = 1)

- No issue with USA players (MSI = 1.0)
- AUS players cleared with lower MSI (.88)





THE ISSUE OF ‘CHEATING’

How is it possible to clear Level 5 (to next level) without getting ESI of 1?

- Perform game replay from Course of Actions (those who clear Level 5 without ESI of 1)
- Found ‘problem’, discuss with collaborator
  - Programmer ‘copy/paste codes’ – adding tiles when doing a particular rotation

No subject reported the ‘issue’ (more tiles) because they can clear the level

- If tiles were taken away (they will probably report)
  - In addition, some probably spread the ‘secret’ in class

Inherent problem with COTS games: assume no bugs (no analytics)

- Problem not discoverable without big data (classroom research not big enough)

‘Cheating’ IS an issue that need addressing in Serious Games Assessment

- Analytics (e.g., ESI) is a viable way to examine this issue

SUMMARY

ESI, SERIOUS GAMES, ORGANIZATIONAL SCIENCE

Analytics research directions (SG assessment, organizational science):

1. Categorization (Clustering Analysis)
2. Prediction (Predictive Analytics)
3. Data Visualizations, Patterns Discovery, ...

As a performance index, ESI is highly functional in:

- Being incorporated into (serious) game analytics – machine learning, data science, R, etc.
- Creating trainee profiles (e.g., scoring, ranking, monitoring individual growth potentials)
- Identifying patterns & abnormally – cheatings, bugs, better solutions, change of strategy, …
- Estimating the cost-benefit ratio of training → better managing organizational performance
- Determining prescriptions for training – the amount and frequency of training

ESI – viable 'performance index' for Serious Games research (Organizational Science)

MULTI-ROUND SERIOUS GAMES DESIGN

ESI for both *ad hoc* (PPI) and *post hoc* (KPI) analysis

- *ad hoc* analysis (PPI) – improve the learning process, identification of talents/problems without waiting for the completion of the training

Advantage: Design games to include multi-round events for assessment

- Eliminate pretest/posttest method (Black Box approach) using 'Repeated Measures' statistics
- Move assessment from 'outcome-based' (summative, KPI) to 'process-based' (formative, PPI)
- Performance comparison – vs self, group, or expert(s)
- Visualization of metacognition: e.g., strategy change shown as 'dips' in the Learning Curve
- Outliers useful for the identification of cheating, software bugs, or discovery of better procedures (even better than expert)
- 'Repeated' data needed for 'Predictive Analytics'

THANK YOU!

csloh@siu.edu