cogent education

EDUCATIONAL PSYCHOLOGY & COUNSELLING | RESEARCH ARTICLE

Learning in balance: Using oscillatory EEG biomarkers of attention, motivation and vigilance to interpret game-based learning

Benjamin Cowley and Niklas Ravaja

Cogent Education (2014), 1: 962236

cogent • oa

CrossMark

Cowley & Ravaja, *Cogent Education* (2014), 1: 962236
http://dx.doi.org/10.1080/2331186X.2014.962236

cogent education

EDUCATIONAL PSYCHOLOGY & COUNSELLING | RESEARCH ARTICLE

Learning in balance: Using oscillatory EEG biomarkers of attention, motivation and vigilance to interpret game-based learning

Benjamin Cowley[1,2,3*] and Niklas Ravaja[3,4,5]

Received: 21 April 2014
Accepted: 01 September 2014

*Corresponding author: Benjamin Cowley, Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, P.O. Box 9, 00014 Helsinki, Finland; BrainWork Research Centre, Finnish Institute of Occupational Health, Helsinki, Finland; EMOID Lab, School of Business, Aalto University, P.O. Box 21220, 00076-AALTO Helsinki, Finland
E-mail: ben.cowley@helsinki.fi

Reviewing editor:
May Cheng, Hong Kong Institute of Education, Hong Kong

Additional article information is available at the end of the article

Abstract: Motivated by the link between play and learning, proposed in literature to have a neurobiological basis, we study the electroencephalogram and associated psychophysiology of "learning game" players. Forty-five players were tested for topic comprehension by a questionnaire administered before and after solo playing of the game *Peacemaker* (Impact Games 2007), during which electroencephalography and other physiological signals were measured. Play lasted for one hour, with a break at half time. We used the Bloom taxonomy to distinguish levels of difficulty in demonstrated learning—with the first five levels assigned to fixed questions—and "gain" scores to measure actual value of demonstrated learning. We present the analysis of the physiological signals recorded during game play and their relationship to learning scores. Main effects related to biomarkers of vigilance and motivation—including decreased delta power and relatively balanced fronto-hemispheric alpha power—predicted learning at the analysed Bloom levels. Results suggest multiple physiological dispositions that support on-task learning styles, and highlight the utility of the psychophysiological method for interpreting game-based learning evaluations.

Subjects: Adult Educational and Lifelong Learning, Behavioral Neuroscience, Continuing Professional Development, Open & Distance Education and eLearning, Psychological Science

Keywords: game-based learning, psychophysiology, electroencephalography, motivation, attention, vigilance, evaluation, serious game

ABOUT THE AUTHOR

Benjamin Cowley obtained BA and PhD degrees respectively from Trinity College Dublin and the University of Ulster, Northern Ireland. Initial post-doctoral projects studied psychophysiology and game-based learning at Aalto University, Helsinki; and attention disorders at the University of Helsinki's Cognitive Brain Research Unit. Current research interests are learning in people and machines, attention and cognition.

The presented study shows the use of EEG to examine the tonic learning of serious game players. This supports the authors' current project, "Revolution of Knowledge Work", where it is asked how novel interactions and the psychophysiological method can be used to improve the symbiosis between knowledge worker and IT system.

PUBLIC INTEREST STATEMENT

We present a neuro-cognitive study of the way that people learn during game play. The electroencephalogram of players was measured while playing *Peacemaker* (Impact Games 2007), a serious game designed to teach a balanced view of the Israel–Palestine conflict. The brain imaging data was analysed, and related to the learning outcomes measured by a knowledge test presented before and after playing. The results show that there are multiple physiological dispositions that support on-task learning styles. Dealing with a complex stimulus environment such as this game, the most successful strategy seems to be one of balance: of balance between the brain's hemispheres and between activation and dissociation.

© 2014 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 3.0 license.

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Downloaded by [213.27.242.194] at 02:00 06 October 2014

1. Introduction

Recent developments in technology enhanced learning (TEL) have opened new formats where students can "learn by doing" (Squire, 2006) in a virtual environment: defined as a *serious game*. However, despite the growing consensus (see background) that such games *can* provide learning given the correct marriage of game and pedagogical design, the field still lacks some key lines of evidence as to *how* this type of learning happens. Such evidence includes the learners' subjective and objective experience and its relationship to learning; the exploration of this relationship using psychophysiological methods is our main motivation.

In the fields of game research (Cowley, Charles, Black, & Hickey, 2008), learning theory (Gee, 2003) and even animal behaviour (Groos, 1898), it has been observed that learning and play are intrinsically (though not necessarily) linked. The type of learning to occur depends on the form of play. One productive line of research has suggested that *action video games* promote improved basic cognitive functioning (Bavelier, Green, Pouget, & Schrater, 2012), facilitating faster learning in general. Game-based pedagogy can also induce other forms of learning. In his analysis of mammalian functional neurology, Panksepp (1998, p. 294) suggests that we possess dedicated and spontaneously developing "play neurocircuitry", with many non-social functions connected to forms of learning: play increases physical fitness, skilful tool use, and the ability to innovate and think creatively. The latter ability, like other higher cognitive abilities such as social skills, is particularly valued in adult and corporate training. Training such soft skills requires a different style of learning game, and research on the physiology of learning in such games is much less well served than the study of action games' effect on basic cognitive functions.

To study the psychophysiological underpinning of learning that occurs in such "soft skills" or knowledge-focused games, we must examine learners' *tonic* brain activity, not only event-related activity as is usually studied. We present results from an experiment designed to address this need, motivated by the opportunity to build greater understanding of *how* serious game play can induce learning. We focus mainly on the electroencephalogram (EEG), notably frequency band power and asymmetry, which relate to such important aspects of learning as vigilance and motivation.

Our results suggest that tonic EEG measures of oscillatory brain waves may serve as a predictor of topic learning. The paper also contributes to the evaluation of game-based learning (GBL) and TEL, and the growing field of applied neuroscience in education and learning, by demonstrating the application of psychophysiological methodology.

In the next section, we describe the relevant state of the art on educational games and measurement of players by psychophysiological methods. The Methods section then details the study methodology, including sub-sections to outline the experiment procedure, the relevant aspects of the test game, the assessment questions and our chosen psychophysiological methods. Section 4 details our results; Section 5 discusses our interpretation and future directions, and Section 6 presents the final conclusions.

2. Background and state of the art

It has been claimed that learning is almost always part of play (Sutton-Smith, 1997). Games generally involve skills (even games of chance can be played more or less skilfully by odds recognition) and Koster (2005), among others, claims that building repertoires of nested skills is the heart of game play progression. Indeed, it is known that skill learning literally has a transformative effect on the player (Scholz, Klein, Behrens, & Johansen-Berg, 2009). Thus, if the delivery is good enough and the games are effective, the serious game player stands to gain a lot.

The efficacy of educational games has been debated and supported in studies for over three decades (Egenfeldt-Nielsen, 2006; Guillén-Nieto & Aleson-Carbonell, 2012; Malone, 1981; O'Neil, Wainess, & Baker, 2005). Kirschner and Clark (2006) claim that discovery, problem-based, experiential

Downloaded by [213.27.242.194] at 02:00 06 October 2014

and enquiry-based techniques are the main tools of games. Habgood and Ainsworth (2011) argue that *intrinsic* motivation is required to make effective serious games. Sung and Hwang (2013)'s study supports the value of collaborative learning in games.

Research to develop a supportive theory for general media and psychophysiology has advanced in recent years, partly due to the reduced cost and improved reliability of psychophysiological measurement equipment. This advance in theory and methods has reflected in a small but growing number of studies on the psychophysiology of learning in serious games (Cowley, Fantato, Jennett, Ruskov, & Ravaja, 2014; Cowley, Ravaja, & Heikura, 2013; Pope & Palsson, 2001; Wamsley, Tucker, Payne, & Stickgold, 2010; Wang, Sourina, & Nguyen, 2011).

2.1. Psychophysiological methods

The psychophysiological method views the mind as being more comprehensible if its physical substrate is considered, in structural and functional terms (Cacioppo, Tassinary, & Berntson, 2000). The method involves using physiological signals, such as scalp potentials, respiration or electrodermal activity, to study psychological phenomena including frustration, mental stress/cognitive overload, approach/withdrawal motivation and attentional processes (Harmon-Jones, 2003; van Oyen Witvliet & Vrana, 1995). The value of psychophysiological methods for TEL evaluation is that the participant cannot give deliberately inaccurate physical signals and the acquisition of signals is non-intrusive, freeing the participant's attention onto the TEL task. Psychophysiology-enabled evaluation can then take place alongside other forms, such as self-report, entirely consistently. These attributes can potentially improve on-task attention during the protocol and reduce the measurement impact of participants' reactivity to being observed (e.g. often termed the "Hawthorne effect" in experiment or clinical trial settings).

A thorough review of psychophysiological methods for game-based experiments can be found in Kivikangas et al. (2011). As suggested there, psychophysiological measurements provide an innovative method for assessing player experiences by indexing emotional, motivational and cognitive responses to entertainment, education, therapy or other types of games (Mandryk & Atkins, 2007; Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006). We next explain how, from existing literature, we derived a psychophysiological approach to studying GBL.

2.2. Psychophysiology and GBL

Tonic values of psychophysiological signals can be used to index various cognitive and emotional processes that can contribute to learning. For example, Chaouachi, Jraidi, and Frasson (2011) examined how EEG recordings obtained during learning tasks could index various aspects of a learner's state. We use tonic signals to fit the type of GBL we study: because learning in this type of game happens over long time periods, with players who can construct concepts from non-linear relationships in the data they are presented with, a more straightforward event-related approach to signal analysis would be less appropriate.

Learning in a broad sense requires environmentally prompted adaptation between states of sustained, focused attention and reflection and internalisation (Clark & Harrelson, 2002). More specifically, EEG should show changes in its feature profile dependent on this adaptation. Based on this, the variation in learning performance across a group can be modelled by features of individual EEG, which will predict the outcome if any group level causal relationship exists between brain oscillations and learning.

Features of the tonic EEG can index attention (Klimesch, 1999), motivation (Harmon-Jones, 2003) or provide a neurophysiological model of vigilance (Roth, 1961). Together with peripheral signals, these indices provide a descriptive set of independent variables (IVs) relevant to any dependent measure of learning. Such EEG features are often derived from a measure of power in (one or more of) the five classic EEG frequency bands— delta δ (.5–3.5 Hz), theta θ (3.5–8 Hz), alpha α (8–13 Hz), beta β (13–30 Hz) and gamma γ (30–45 Hz).

Downloaded by [213.27.242.194] at 02:00 06 October 2014

2.2.1. Cognitive performance

All frequency bands are also individually linked to signs of performance (i.e. learning prerequisites). For example, it has been observed that posterior alpha desynchronisation accompanies cognitive tasks (Klimesch, 1999).

Similarly, power in the β and γ bands is well known to vary in relation to task demands (Palomäki, Kivikangas, Alafuzoff, Hakala, & Krause, 2012)—indeed event-related synchronisation of higher frequency ranges of the EEG can be a powerful tool for analysing cognitive processing (Krause, 2000). In prior work, β band power has been associated with phase-synchronisation of remote areas of attention networks (Gross et al., 2004), while the ratio of β–θ power has been suggested as an index of task engagement (Kamzanova, 2011). The conceptualisation of γ has been suggested as a selectively distributed parallel processing gamma system (Başar, Başar-Eroğlu, Karakaş, & Schürmann, 1999), representing a universal code of central nervous system communication.

Other results suggest that δ power can be an active component during learning. Mathewson et al. (2012) show that for a "complex video game" (*Space Fortress*), δ activity from 250 to 600 ms after an important event was positively associated with game-score indexed learning rate. Karakaş and colleagues stated that δ is an integral component of task-relevant responding, "Delta response thus represents cognitive efforts that involve stimulus-matching and decision with respect to the response to be made" (Karakaş, Erzengin, & Başar, 2000, p. 48). Yet, because these findings are related to event-related paradigms, their applicability to tonic-data analysis remains unclear.

In the work on attention deficit disorder, tonic δ wave power has been associated with inattentive states (Markovska-Simoska & Pop-Jordanova, 2010). A combined functional magnetic resonance imaging and EEG study (Jann, Kottlow, Dierks, Boesch, & Koenig, 2010) has shown that all the resting state networks associated with higher cognitive functions such as self-reflection, working memory and language all displayed a positive association with higher EEG frequency bands, while negatively related to delta and theta. Knyazev's literature review on δ oscillations gives an explanation from an evolutionary perspective (Knyazev, 2012).

2.2.2. Vigilance

Further features available from EEG band powers include the vigilance model (Roth, 1961). Vigilance states with established EEG indices range from relaxed wakefulness, marked by posterior α, to sleep onset marked by the occurrence of K-complexes and sleep spindles. Low-voltage EEG, meaning increased δ and θ activity, is observed as vigilance wavers between low and wakeful and thus provides an indicator of baseline likelihood of task engagement (Minkwitz et al., 2011). Vigilance regulation, maintaining a task-appropriate level of attention and arousal, is the core feature of learning. It is worth noting that vigilance is not simply a scale of activation from awake to asleep, but also the readiness to deploy directed attention, which may change levels while the individual remains at the same level of arousal.

2.2.3. Motivation and arousal

Another EEG feature derived from band power is hemispheric asymmetry. The asymmetry between left and right fronto-hemispheric α power may signify motivational states, according to the model of Davidson and others (Harmon-Jones, 2003). Relatively greater right frontal activation is associated with withdrawal motivation, and relatively greater left frontal activation with approach motivation. Source localisation of frontal asymmetry in the alpha frequency band (i.e. the index of frontal asymmetry in EEG studies) has indicated that it reflects activity in the dorsal prefrontal cortex (Pizzagalli, Sherwood, Henriques, & Davidson, 2005). This area is primarily known for integration of sensory and mnemonic information and the regulation of intellectual function and action, which are key aspects of conceptual learning.

An optimal arousal level has been proposed to facilitate learning (Baldi & Bucherelli, 2005; Sage & Bennett, 1973), and indeed it is important to contextualise EEG signals by the arousal level of the individual. Arousal is most often measured with EDA (or skin conductance level; also sometimes

Downloaded by [213.27.242.194] at 02:00 06 October 2014

called galvanic skin response) (Bradley, 2000; Dawson, Schell, & Filion, 2000), so EDA is an often-used physiological measure for studying digital gaming experiences (Mandryk & Atkins, 2007; Staude-Müller, Bliesener, & Luthman, 2008). The neural control of eccrine sweat glands—the basis of EDA—predominantly belongs to the sympathetic nervous system that non-consciously regulates the mobilisation of the human body for action (Dawson et al., 2000).

2.3. Hypotheses

Following these models of band powers (Jann et al., 2010; Minkwitz et al., 2011), our first hypothesis is

H1: lower vigilance/task engagement as indexed by relatively greater low-frequency (delta or theta band) EEG activity will predict worse learning performance as indexed by assessed scores from pre- to post-learning tests.

We consider that those in a low vigilance state should not evince approach motivation, so that we propose hypothesis

H2: poor learning performance will be predicted by relatively higher right frontal hemisphere asymmetry accompanied by increased low-frequency EEG.

Additionally, since approach motivation in the context of learning suggests the probability of task-related synchronisation (i.e. deployment of neural resources), we propose

H2a: high learning performance will be predicted by relatively greater left frontal asymmetry especially when beta or gamma synchronisation is high.

The physiological activation entrained by this neural activation should also show in participants' arousal, so we propose

H2b: high learning performance will be predicted by relatively greater left frontal asymmetry especially when EDA is high.

3. Methods

3.1. Design

In our experiment, we wished to relate tonic physiological data to the learning outcomes of serious game players. Participants were recruited to play one hour of the Peacemaker serious game (Impact Games 2007), which aims to teach the player about the nature and causes of the Israel–Palestine conflict and has been quite successful (Burak, Keylor, & Sweeney, 2005). We tested learning outcomes using questionnaires delivered before and after play, and analysed these outcomes with respect to the psychophysiological state of the learner during play. With this approach, we aim to track the interplay between the players' physiology and the learning outcomes from GBL. Assessment of learning was controlled by splitting participants into two conditions, where the second condition had a mid-play period of discursive reflection in groups of two–three. Differences observed between groups demonstrate that their learning outcomes were not simply the result of test repetition, but that the inter-test intervention of game playing had an effect.

3.2. Participants

Recruitment of participants was conducted by advertising the study over student internet mailing lists. Potential volunteers were asked to respond "yes" or "no" if they had some prior exposure to the topic of the learning game: personal connections to Israel or Palestine or significant prior knowledge of the subject matter. These responses were used as exclusion criteria to prevent bias in the learning process.

A total of 45 participants (16 females, 29 males) volunteered in exchange for non-remunerable department store vouchers. Of the 45 participants, data-sets from 10 were excluded during the analysis due to corruption of the EEG data by artefacts, so that the final sample was 35 (15 females). In accordance with the declaration of Helsinki, participants were thoroughly briefed on the purpose

Downloaded by [213.27.242.194] at 02:00 06 October 2014

and procedure of the study; each signed a written informed consent prior to the experiment. Participants were also reminded that they could withdraw from the study at any time without fearing negative consequences. As the study did not concern medical research, it required, in accordance with Finnish law, no formal ethical approval from the Ethics Review Board of Aalto University. Before testing, extra background information was obtained by means of a short questionnaire. Participants were mostly Finnish students or graduates, all non-native English speakers aged from 19 to 32 years (mean M = 24.7, standard deviation SD = 3.6), and had an average level of computer-game playing frequency (on a scale of "1: Not a lot" to "5: A lot", M = 3, SD = 1).

3.3. Procedure

The experiment procedure was divided into six main phases, as shown in Figure 1. First, participants answered 41 questions concerning the Israel–Palestine crisis, which took an hour (M = 56.2 min, SD = 18.7)—a time that did not significantly vary between conditions ($t(43)$ = .85, *ns*.).

The second phase consisted of attachment of psychophysiological sensors (see details below). Each participant was seated in an electrically shielded laboratory for impedance inspection and game-play. This process took, on average, 72 min (SD = 32).

Next in phase 3, the participants were seated in front of computers and played a game tutorial (M = 7.4 min, SD = 1.7) and the first of two 30-min gaming sessions. For condition 2, participants played alone, physically and in the game.

The two game sessions were broken by phase 4. For condition 1, this consisted only of answering two quick experiential self-report questionnaires on mood and performance (not analysed herein). Condition 2 differed from condition 1 by the presence of a reflection period during phase 4: the players were brought into a group to participate in a guided discourse reflecting on their game experience, in addition to completing the self-reports. This discussion was the only point at which participants in condition 2 were not visually and aurally isolated from each other, so as to create a similar *playing* experience in both conditions. The lead experimenter directed the discussion period, so that it remained on topic, encouraging free discussion.

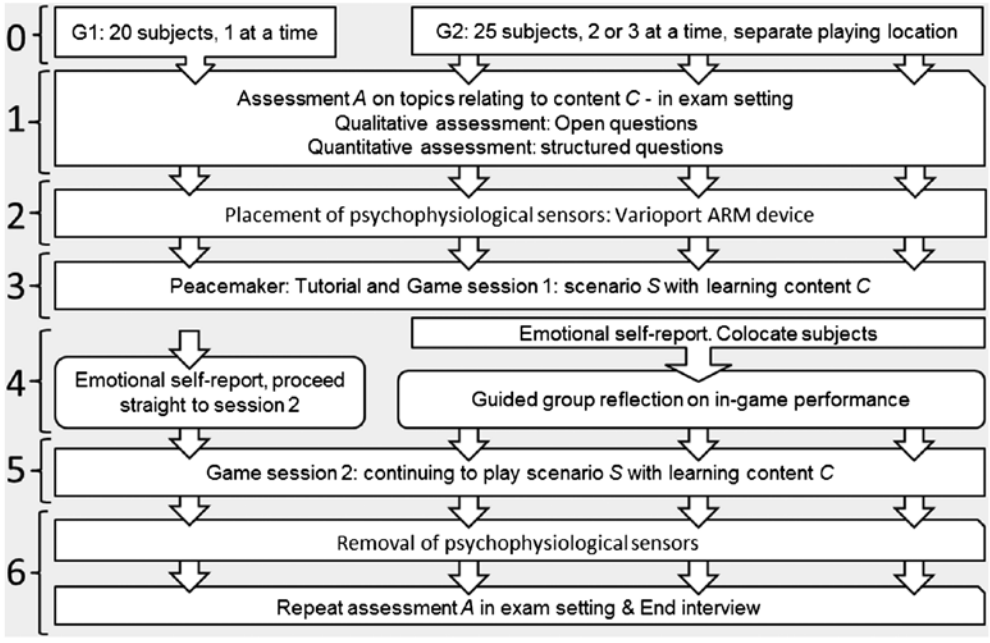

Figure 1. The experiment protocol setting out activities for two groups.

Downloaded by [213.27.242.194] at 02:00 06 October 2014

In phase 5, the second 30-min game session was played. The monitoring equipment was removed, and total time attached to electrodes was M = 102 min, SD = 12. The sixth and final phase of the experiment was to answer the 41 questions a second time, taking on average 33.7 min (SD = 12.6) again without significant difference in time taken ($t(43) = .95$, *ns*.).

3.4. Proxy game

The *Peacemaker* serious game, shown in Figure 2, was designed to teach a peace-oriented perspective on the Israel–Palestine conflict. For a thorough study on the interaction effects between psycho-social personalities of players and their performance in Peacemaker, see Gonzalez and Czlonka (2010). It is a point-and-click strategy game, where the player acts as a regional leader and must choose how to react to the (deteriorating) situation, deploying more or less peaceful options from diplomacy and cultural outreach to police and military intervention.

Play is oriented around strategic management of conflict, taking governmental actions as shown by the menu on the left. Conflict is modelled by factions/stakeholders who each have approval ratings for the player—information can be obtained by clicking on a faction's icon. "Spontaneous" events are reported as news (marked on the screenshot by reticules), which drive the game narrative, and as player approval ratings with a particular faction vary, these events become more or less critical (in the screenshot, crisis is indicated by the colour of the reticule). Events and player actions are combined to drive approval ratings—winning is defined as achieving 100/100 on both Israeli and Palestinian ratings (see bottom left), while losing happens after scoring −50/100 on either.

Thus, players are expected to learn a new and more subtle perspective on the Israel–Palestine situation, as well as insights into the requirements of stakeholder management in a potential conflict scenario, and the capacity for dynamic decision-making (Gonzalez & Czlonka, 2010). The Peacemaker game supports these requirements; in fact, its benefit as a learning tool has caused it to be

Figure 2. Screenshot of the Peacemaker game.

Downloaded by [213.27.242.194] at 02:00 06 October 2014

internationally used.[1] Thus, the fit to the TARGET requirements was good: Peacemaker may be played in a short duration without extensive pre-training, and imparts valuable insights into conflict resolution even in a short duration.

3.5. Questions and assessment

To assess learning, we chose a pre–post-test design using questionnaires with quantifiable accuracy. Certain criteria apply to such designs. The questions must be answered pre-game, so they could not reference too specifically the content in the game, but must be answered again post-game and also be able to elicit the participant's learning of the topic represented by that content. The questions also need to address all the (Bloom) levels of learning which the game provides scope for. The Bloom taxonomy of learning levels (Anderson, Krathwohl, & Bloom, 2001) describes the difficulty of attaining a particular level of learning—the levels themselves being represented (in Bloom's system) by descriptions of the kinds of content one would produce to show attainment of such learning.

The 47 questions (including four open questions) were generated by the authors mining the content of the game (accessed from the spreadsheets that store the textual game content). Questions were thus all designed to tap the knowledge which could be learnt from the game, and constrained to be valid by the method of sourcing from the game material. We assigned Bloom levels based on complexity of interactions between content in the question itself and the acceptable answers to the question. For instance, first-order interactions exist between a question such as "What is the religious capital of Israel?" and the answer "Jerusalem", which would place this question at the first Bloom level.

In the Appendix, Table A1 outlines the relationship between types of questions, the Bloom level assigned to them and the game data or experience which the question addresses—it also lists the number of such questions asked. Also in the Appendix are a sample list of questions and details of the assessment protocols, for readers with greater interest in the educational aspect of the study. Here, we describe the assessment sufficiently to understand the DV used in the analysis. Assessment protocols were developed for each Bloom level; for level 1–5, the protocols gave comparable scores and were thus combined to a final learning score, while a separate score was derived for Bloom level 6 open questions.

Open questions required a more qualitative approach, whose final quantification was not comparable on an interval scale to the level 1–5 questions. Unfortunately, the level 6 results did not have very high variance, since the majority of participants could not be considered to have demonstrated this high level of learning in their answers and therefore had a level 6 score of zero. The inter-rater reliability for the level 6 questions was also not good, mostly less than .4 "poor to fair agreement"; thus, level 6 results are not included in the analysis below.

For the first five Bloom levels, we derived a "correct" answer from the game documentation and data mining of empirical records (logs) of games played—i.e. a "truth" value in relation to each question was established by studying *what the game had shown the players*. Using these answers, we assessed fixed-choice responses by scoring the difference between the subject's first and second response with respect to how much more accurate (or inaccurate) they became, i.e. *gain scores*. Normalised gain scores were considered a non-prejudicial approach with high flexibility, in that the gain could be readily transformed for weighting or data exploration, as advised by Lord, French, and Crow (2009, p. 22).

Before summation to a final learning score, gain scores for each question were weighted by the Bloom level rating of the question associated (giving more weight to questions that theoretically indicated a higher level of learning) and then normalised. We used a non-linear weighting scheme to reflect the relatively greater importance of higher levels of Bloom learning; for example, higher level learning can be considered of parametrically greater importance than lower level learning, because mastery at each level requires mastery at all the lower levels first. Thus, the weights [1, 2, 4, 8, 16] were applied to levels 1–5 (for a rationale on linear vs. non-linear weighting, see Gribble, Meyer, and Jones [2003]).

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Thus, our final dependent variable was the exponentially weighted normalised gain score based on level 1–5 questions, which we will refer to as *gainExp* (M = 11.68, SD = 13.02, range [−14, 42]). The *gainExp* distribution was normal by Kolmogorov–Smirnov test (with Lilliefors Significance Correction), $D(35) = .08$, *ns*.

3.6. Psychophysiological data acquisition and pre-processing

For data acquisition, we used the Varioport-ARM multi-amplifier biosignal-recording device (Becker Meditech). We recorded the psychophysiological signals EEG, ElectroOculogram (EOG), EDA and respiration. EEG was recording from six Ag/AgCl electrodes on a cloth cap following the 10–20 system (Niedermeyer, 2005, p. 140) at F3, F4, C3, C4, P3 and P4. AFz was used as ground and the reference montage was linked to ear clips. For eye-blink/saccade artefact correction, EOG was recorded by bipolar Ag/AgCl electrodes placed ~2 cm above and below the left eye for vertical saccade, and ~1 cm from the outer canthi for horizontal saccade. EDA was recorded from the proximal phalanges of the index and middle fingers of the non-dominant hand. Respiration was recorded using an adjustable belt transducer placed around the chest. All channels were recorded at a sampling rate of 1000 Hz and down-sampled online where appropriate. Impedance testing was carried out to ensure less than 5 kΩ resistance, and 8 min of baseline were recorded. For pre-processing, Variograf software was used to "read and reconstruct" binary data into *vpd* format files from which separate software was used for each signal.

EDA signal was pre-processed using the Ledalab (v 3.43) toolbox for Matlab 2010b in batch mode: signal was down-sampled to 16 Hz and filtered using Butterworth low-pass filter with cut-off 5 Hz and order eight. Then, the signal was divided into phasic and tonic components using the non-negative de-convolution method (Benedek & Kaernbach, 2010).

For EEG analysis, Brain Vision Analyser v1.05.005 (BVA) was used to pre-process the *vpd* files in eight steps. We first applied Butterworth zero-phase filters, with time constant .3 s and 12 dB/octave roll off, at high pass of .5 Hz and notch of 50 Hz. Second, pulse artefacts from heart rate interference were detected and corrected using BVA's MRI algorithm, taking the R-peak latency from the ECG channel with an average over 10 pulse intervals. The third step was ocular artefact correction using Gratton and Coles' algorithm (Gratton, Coles, & Donchin, 1983) with input from EOG channels commonly referenced with the EEG. The fourth step was segmentation into 1-s epochs (extracted from the trials of interest), followed fifth by BVA's Artefact Rejection algorithm testing 100 ms intervals for minimum/maximum amplitude of ±200 MV and lowest allowed activity (maximum minus minimum) of .5 MV. Sixth was Fast Fourier Transform power density calculation over 1-s epochs, with 10% Hanning window and resolution .5 Hz. In the seventh step, we corrected for myogenic noise, i.e. artefacts from gross motor interference by the participant, including jaw clenching and head scratching. Due to the low channel count, blind source separation methods were unsuitable for this correction—instead the power regression method was used, which Davidson described initially and again validated more recently with others (McMenamin, Shackman, Maxwell, Greischar, & Davidson, 2009). The regression method was implemented in Brain Analyser v1.05 by the authors, and compares power density between the alpha band and the high-frequency band 70–85 Hz. Finally, the eighth step was feature selection, described below.

- Power in the five EEG frequency bands δ, θ, α, β and γ was obtained from the mean of the six recording electrodes and band-pass filtered with settings as described above.
- Frontal (F) asymmetry of EEG was derived by taking the natural logarithm of the product of mean alpha power in F3 and with the reciprocal of F4, that is, $\ln(\alpha{:}F3 \div \alpha{:}F4)$. With odd numbered electrodes on the left-hand side of the head, this equation implies that relatively greater left asymmetry is denoted by positive numbers (i.e. $\alpha{:}F3 > \alpha{:}F4 \geq \alpha{:}F3 \div \alpha{:}F4 > 1$, and $\ln([1, \infty)) \rightarrow N$).
- Tonic EDA was obtained by the NND method as explained above.

Downloaded by [213.27.242.194] at 02:00 06 October 2014

3.7. Statistical analysis

To obtain IVs for statistical modelling, mean values of each feature were derived from 1 min epochs across the playing periods, giving a data-set with 60 rows per participant and one column per IV, DV or factor. This "tonic" approach allowed us to test for relationships that hold across trial duration but are not specific to individual, potentially non-repeating events. Data was examined to check for distribution characteristics. To achieve approximate normality, we rectified the data with an additional constant to achieve minimum bound of 1.0 and calculated the *z*-scores. After excluding any rows that had a *z*-score greater than 2.58 (i.e. any outliers plus the most extreme 1% of the distribution), data was transformed by taking the square root. With all data ≥1.0, this transform preserved relative values while helping to correct skew. Although the data was still not normal by Kolmogorov–Smirnov tests, this was not unusual for large data sets according to Field (2009, p. 139), whose visual criterion (histogram-to-normal curve matching) and *z*-score criterion (95% < 1.96, less than 1% > 2.58) were used to judge that the data showed a good approximation to normal.

The generalised estimating equations (GEE) procedure in SPSS was used to test all hypotheses, to support a repeated measures model over the 60-epoch rows. We specified participant ID as the "Subject" variable and trial number and minute as the within-subject variables. On the basis of the "quasi-likelihood under independence model criterion", we specified autoregressive as the structure of the working correlation matrix. We specified a normal distribution with identity as the link function. DV was *gainExp*, and IVs were epoched features of the physiological signals, as mentioned in the previous section: δ, θ, α, β and γ band power; frontal asymmetry; and tonic EDA.

Due to the natural variation between individual physiologies, psychophysiological data must always be baseline-corrected before analysis. This is done by adding one extra factor to each model for each IV, corresponding to the mean value of pre-play baseline measurement of the signal for that IV. The final factor in the models reported is *Condition*, which was added to all models as a control. Although the analysis resulted in multiple tests, multiple comparison testing was not performed because the comparisons were planned and the IVs based on band powers are not independent.

GEEs are an extension of the generalised linear model, and were first introduced by Liang and Zeger (1986) and Ballinger (2004) for a more complete introduction to GEEs for longitudinal data analysis. GEEs allow relaxation of many of the assumptions of traditional regression methods such as normality and homoscedasticity, and provide the unbiased estimation of population-averaged regression coefficients despite possible misspecification of the correlation structure. Where psychophysiology is modelled in several variables, the usual assumption of independent observations would be violated. Unless the model accounts for the "within" correlation, the result may inflate the Type II error; thus, GEEs suit well the analysis of time series psychophysiological data.

4. Results

The *gainExp* variable had $M = 12$ and SD = 12, ranging from −14 to 42. To illustrate that significant learning did occur, we performed a *t*-test on *gainExp* against a mean of 0, $t(44) = 6.9$, $p < .001$. Furthermore, we compared the learning scores for each condition to show that learning outcomes were not independent of the inter-test intervention. Scores were significantly different between condition 1 and condition 2, by independent samples *t*-test, $t(43) = 2.8$, $p < .01$, with the direction of difference-favouring condition 1. Thus, the group who did not have a mid-play reflection session had higher *gainExp* score; the effect was reported in more detail in Cowley, Heikura, and Ravaja (2013).

The statistically significant psychophysiological results are summarised in Table 1, showing each physiological variable (IV) that predicted learning (DV) and associated statistics.

To explore the relationship between EEG band power and learning, we modelled *gainExp* scores with band power, by specifying one GEE for every EEG band. Covariates were the main effects of baseline band power and task-level band power (power during game play). Supporting H1, task-level

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Table 1. Summarised statistical results, only significant results are reported

IVs	*B*	SE	Wald $\chi^2_{(df=1)}$	DV
δ-band power**	−.001	.0003	3.9	*gainExp*
F-asymmetry × β-band power**	.001	.0004	5.4	*gainExp*
F-asymmetry × γ-band power**	.0001	.0001	4.7	*gainExp*
F-asymmetry × EDA*	.001	.0003	3.9	*gainExp*

*$p = .05$.
**$p < .05$.

δ-band power was negatively associated with *gainExp* scores, $B = -.001$, SE = .0003, Wald $\chi^2_{(df=1)} = 3.9$, $p < .05$. Thus, a potential indicator of reduced attentiveness and vigilance tended to increase as learning performance decreased.

The relationship in our sample between learning and power density in each EEG band is shown in Figure 3 below, where the top row is the grand average scalp-distributed power density of high-scoring players (median split on *gainExp*) and the bottom row is the grand average of low-scoring players. Grand averages were derived from regression-corrected data in BVA. One can clearly see the difference between scoring levels, especially in δ as high scorers have low frontal power and low scorers have high frontal power.

Moving to the construct *motivation*, the claim of H2a *was* supported for *both* EEG bands: to wit, that greater relative left frontal asymmetry accompanied by β or γ synchronisation would predict *higher* learning scores. F-asymmetry × β-band power significantly predicted *gainExp*, $B = .001$, SE = .0004, Wald $\chi^2_{(df=1)} = 5.4$, $p < .05$, while F-asymmetry × γ-band power also predicted *gainExp*, $B = .000$, SE = .0001, Wald $\chi^2_{(df=1)} = 4.7$, $p < .05$. Each of these results used a separate GEE model with main effects of baseline F-asymmetry, baseline band power, task-level F-asymmetry, task-level band power; all as covariates; and the task-level F-asymmetry × task-level band power interaction.

To explore the role β and γ play in the interaction, we made a graphical examination of the levels of the interactions, see Figure 4. Each panel displays two levels of F-asymmetry on the abscissa, spilt at quartile 1; mean *gainExp* is on the ordinate; two levels of each IV are depicted by the ◊ (diamond) and ● (ball) symbols. β is split at the median and γ is split at quartile 1. Thus, Figure 4 panels A and B show that the effect of motivation on learning scores, for instance approach motivation indexed by

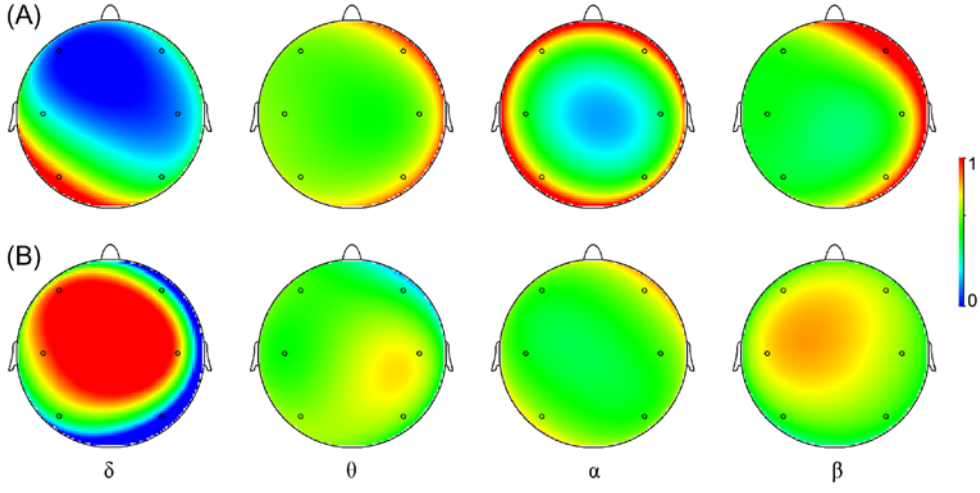

Figure 3. Head maps of band power density averaged over game sessions, for δ, θ, α, β.

Notes: Row A shows high-scoring players (by median split); Row B shows low-scoring players. Scale is normalised between zero and one.

Downloaded by [213.27.242.194] at 02:00 06 October 2014

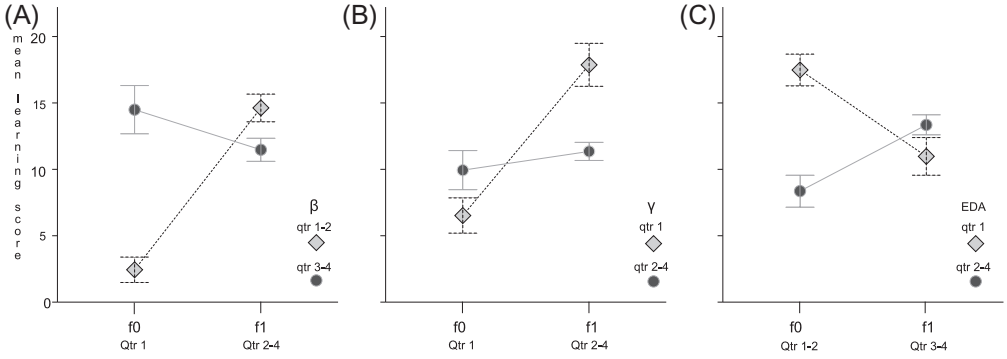

Figure 4. Three interactions involving F-asymmetry shown in panels: A–β, B–γ and C–EDA.

Notes: The mean of *gainExp* is on the ordinate, and error bars are 95% CI. F-asymmetry is split at quartile 1 in panels A and B; it is median split in panel C. β is median split; γ and EDA are split at quartile 1. Lower and upper split portions are shown by ◊ and ● symbols, respectively.

relatively greater left frontal asymmetry, may be modulated by task-related neural synchronisation. We can also see this in Figure 3, where high-scorers also show greater right frontal beta power than low scorers, mirroring Figure 4 panel A.

Finally, H2b claimed that motivation is a natural concomitant of physiological arousal, indexed by EDA. A link between them and learning performance was supported by the result of a GEE model with main effects of baseline F-asymmetry, baseline EDA, task-level F-asymmetry, task-level EDA; all as covariates; and the task-level F-asymmetry × task-level EDA interaction, which predicted *gainExp* with marginal significance, $B = .001$, SE = .0003, Wald $\chi^2_{(df = 1)} = 3.9$, $p = .05$. To examine this interaction, again we performed visual analysis of the variables and selected the most informative panel to display in Figure 4, panel C, where the abscissa shows the median split of F-asymmetry and EDA is split at quartile 1. Panel C shows that the effect of frontal asymmetry was modulated by the relative arousal of the participants.

5. Discussion

In this study, we examined how psychophysiological indices of attention, arousal, vigilance and motivation during playing of a serious game help to clarify the players' likelihood to learn declarative knowledge.

The δ vs. *gainExp* result is novel for the type of learning measured, but in terms of interpretation, the role of δ oscillations described in the literature is not at all simple. However, it is relevant that results, which suggest that δ is linked to learning, have come from event-related analyses, whereas tonic studies of δ waves such as ours have tended to suggest that excess δ is a sign of inattention (Markovska-Simoska & Pop-Jordanova, 2010) or low vigilance (Minkwitz et al., 2011).

5.1. Fronto-hemispheric asymmetry and learning

The interaction of F-asymmetry with both β and γ-band power predicted learning, in both cases with positive relation. The graphical investigations of the relationship between F-asymmetry and *gainExp* showed that it is usually positive: mean learning scores are higher when left frontal power is relatively greater; this is regardless of the value of interacting variables—*except* at the levels shown in panels A and C in Figure 4. These two panels show the modulating influence of certain levels of β and EDA wherein participants with relatively *right* frontal power scored better. These were the only circumstances in which the relationship between F-asymmetry and *gainExp* is negative. The positive relationship is not a main effect (the GEE model F-asymmetry vs. *gainExp* was not significant), but it provides the background against which to consider the three interactions involving F-asymmetry.

Frontal asymmetry is suggested to index motivation, with relatively greater left activation signifying approach motivation and vice versa. Thus also of note is the range of F-asymmetry: it peaks at −.5, suggesting that right frontal power was dominant over left and, *in general,* motivation was more "withdraw" than "approach".

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Figure 4 panel C shows that the overall positive relationship of F-asymmetry vs. *gainExp* is strongly reversed for the lowest quartile of EDA. And in the lower half of F-asymmetry, i.e. stronger withdrawal motivation, the two arousal levels show large differences in learning score. The combination illustrates a group of participants who were probably not well focused: perhaps due to boredom and fidgeting. From another perspective, when highly aroused it paid to be more approach motivated; when less aroused, the opposite was true.

Panel A shows F-asymmetry split at the lowest quartile and β split at the median; low β implies low scores when withdrawal motivation is strongest, but for participants with more balanced motivation, their scores with low β band power exceed those with higher power. Similarly for F-asymmetry $\times \gamma$, there is an adjustment of the effect of low band power when F-asymmetry is more balanced. In both cases, this adjustment is more evident when these upper bands contain less power; the effect of F-asymmetry on learning is greater—it pays more to be in the middle "neutral" motivational state. When β or γ contain more power, the hemispheric power distribution is less relevant.

We stated that F-asymmetry was predominantly *right*, indicating withdrawal motivation. The participants' self-reports were generally positive (valence was 12% above neutral; positive affect was 16% above neutral), which warrants a closer look at this issue and at the possible interpretations of F-asymmetry. In Cowley, Ravaja, et al. (2013), we showed how decreased mental workload (i.e. cognitive efficiency) and positive affect predict increased learning (in the same study). Rotenberg and Arshavsky (1997) showed that a mental imagination task can increase right hemisphere activity. Relevant to this is the fact that the Peacemaker game is an abstract simulator—in other words, it simulates a scenario but does not show explicit representations of the actors or events contained there; rather, it evokes these in the player's imagination using icons, news reports and narrative. Gable and Harmon-Jones (2008) state that the intensity of motivation determines the focal range of attention—low-intensity motivation, whether approach or withdrawal, results in broader attention.

Taking all observations into account, we suggest that the F-asymmetry result shows that playing the game induced a more imaginative cognitive approach characterised by greater right hemisphere activation. Figure 4 suggests that the highest learning scores were obtained by those who had either a) low-arousal and -withdrawal motivations or b) reduced high-frequency band power and more balanced motivation. Taking Gable and Harmon-Jones (2008) into account, the latter group b) suggests that lower intensity motivation—and thus broader attention—is positively modulated by lower β and γ, which, as indices of integrative attention networks, might indicate the benefit of reduced distractibility. Meanwhile, the former group a) appears to be a balance between intensity of motivation and level of arousal; especially recalling that those with relatively greater *right* F-asymmetry performed better when their high-frequency band power was higher, this group appears to represent those who were *on-task* and focused. The cognitive efficiency interpretation we earlier proposed supports these explanations.[2]

Prior results on asymmetry mostly arise from classical event-related protocols, contrasting with our experiment. It is natural that if local hemispheric regions support distinct functions, then the more varied are the range of functions in a protocol, the more both hemispheres must be activated (McGilchrist, 2009, p. 26). Thus, we might say: the participants who displayed more task engagement in Peacemaker's continuous information integrating protocol were more likely to use their whole frontal cortex and evidence more balanced mean power.

The asymmetry results also seem to link to the vigilance result because the lowest scorers had the highest withdrawal motivation rating *and* highest delta values, suggesting their withdrawing and lack of vigilance sprang from the same source—perhaps one engendered the other, or both were engendered by dissociative mood.

Downloaded by [213.27.242.194] at 02:00 06 October 2014

5.2. General issues and future work

The results show us that various measures of the physiology can be predictive of learning, as measured by a self-report questionnaire. There are naturally several caveats as follows.

The learning measure itself must be understood as an imperfect and limited measure, because it is not possible to design a reasonable-length questionnaire to cover all things that can be learnt in a serious game. In the light of this, our claims should not be interpreted as over-reaching.

The seven protocol phases described were designed to help achieve a measurable learning outcome. Orientation of the participants to the topic by the pre-test was a concern: the long period of distraction during sensor attachment may have partially addressed it. The game session length was maximised with respect to the overall length of the experiment and the other periods, to enable better chance of learning by prolonged exposure. There was an impetus to minimise the total time of the learning exercises to reduce the discomfort of wearing the sensors. Nevertheless, we used a total playing time which was cohesive with that used in other Peacemaker studies (Gonzalez & Czlonka, 2010), where reasonable learning results were reported.

The complexity of the results, with many interactions, hints that one should not expect a simple linear relationship between learning and a given psychophysiological construct. It may be valid to use a single-trial analysis to look for such relationships, and some evidence suggests that such analyses can cluster events in the game play around stable and significant psychophysiological reactions (Cowley et al., 2013). In future work, the study of these event reactions could give further insights into GBL.

In terms of experiment design, it would be ideal to increase the sample size. $N = 35$ is small compared to most learning studies; however, it is more than the usual sample size for psychophysiological experiments. Since our main focus is on the psychophysiological method, $N = 35$ is sufficient for reporting existing results. It is also apparent to the authors that repeated sessions of the same protocol would permit a more thorough analysis, while changing games every session to avoid practice effects.

EEG was used to characterise and measure attention, with the ATT index and others. However, the proper measurement of attention should include behavioural measure dependent variables. Unfortunately, these could not be explicitly included in our protocol task, as it was dedicated to learning. Nevertheless, we can assume with some confidence that such constructs as attention are included in the final performance scores from the game and questionnaire.

6. Conclusions

We reported on a study the psychophysiological correlates of learning in serious games. The learning test instrument was assessed in its two parts, a set of fixed-format questions and a number of open questions, all on the topic of the game. The significant results apply only to the fixed-format questions, mainly because many participants did not display Bloom level 6 learning.

In summary, we found that participants who displayed less δ-band power and had an elevated RR and ATT index, and those with more balanced F-asymmetry, were more likely to score highly. Some exceptions exist, such as that for the highest levels of RR, it can be beneficial to have increased δ-band power, or that those with low arousal performed better when F-asymmetry was more imbalanced. The implication of these results is that participants' learning styles are sub-served by differential activation patterns of the physiology. It may be useful to consider this result in designing similar games and their pedagogical application.

By dint of the detailed picture they presented, the psychophysiological methods used show their usefulness for experience analysis, which can be considered a bonus in the context of studies in the TEL field—perspectives on this argument from a similar study were also presented in Cowley et al. (2014).

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Acknowledgements
Authors would like to thank Lauri Janhunen who gathered and processed data; Svetlana Kirjanen and Christina M. Krause who gave critical input. Also thanks to Dr Ingmar Gutbertlet for technical assistance with BVA, and Eric Brown for generously supplying Peacemaker logging software.

Funding
This work has been funded by the EU project TARGET [grant number IST 231717].

Author details
Benjamin Cowley[1,2,3]
E-mail: ben.cowley@helsinki.fi
Niklas Ravaja[3,4,5]
E-mail: niklas.ravaja@aalto.fi
[1] Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, P.O. Box 9, 00014 Helsinki, Finland.
[2] BrainWork Research Centre, Finnish Institute of Occupational Health, Helsinki, Finland.
[3] EMOID Lab, School of Business, Aalto University, P.O. Box 21220, 00076-AALTO Helsinki, Finland.
[4] Department of Social Research, University of Helsinki, Helsinki, Finland.
[5] Helsinki Institute for Information Technology HIIT, Finland.

Citation information
Cite this article as: Learning in balance: Using oscillatory EEG biomarkers of attention, motivation and vigilance to interpret game-based learning, B. Cowley & N. Ravaja, *Cogent Education* (2014), 1: 962236.

Notes
1. See for instance http://gaming.wikia.com/wiki/PeaceMaker_(video_game) and also http://phe.rockefeller.edu/docs/PeresCenterPressRelase.pdf.
2. There is an interesting link between these conclusions and the seminal work of Malone (1981), who observed that learning games worked best when evoking "curiosity" and "fantasy".
3. Competing but equally correct answers are not what was initially listed in the game documentation (which gave the original basis for forming the question), but were proven to be equally valid by empirical means (mining the game log files of participants).

References
Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.
Baldi, E., & Bucherelli, C. (2005). The inverted "U-shaped" dose-effect relationships in learning and memory: Modulation of arousal and consolidation. *Nonlinearity in Biology, Toxicology, Medicine, 3*, 9–21. doi:10.2201/nonlin.003.01.002
Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods, 7*, 127–150. doi:10.1177/1094428104263672
Başar, E., Başar-Eroğlu, C., Karakaş, S., & Schürmann, M. (1999). Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG? *Neuroscience Letters, 259*, 165–168. doi:10.1016/S0304-3940(98)00934-3
Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience, 35*, 391–416. doi:10.1146/annurev-neuro-060909-152832
Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*, 647–658. doi:10.1111/j.1469-8986.2009.00972.x
Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 602–642). New York, NY: Cambridge University Press.
Burak, A., Keylor, E., & Sweeney, T. (2005). PeaceMaker: A video game to teach peace. In M. Maybury, O. Stock, & W. Wahlster (Eds.), *Intelligent technologies for interactive entertainment: First International Conference, INTETAIN* (pp. 307–310). Madonna di Campiglio: Springer. doi:10.1007/11590323_40
Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). Psychophysiological science. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 3–26). Cambridge: Cambridge University Press.
Chaouachi, M., Jraidi, I., & Frasson, C. (2011). Modeling mental workload using EEG features for intelligent systems. In D. Hutchison, T. Kanade, J. Kittler, & J. M. Kleinberg (Eds.), *User modeling, adaption and personalization* (Vol. 6787, pp. 50–61). Heidelberg: Springer Berlin. doi:10.1007/978-3-642-22362-4_5
Clark, R., & Harrelson, G. L. (2002). Designing instruction that supports cognitive learning processes. *Journal of Athletic Training, 37*, S152–S159.
Cowley, B., Charles, D., Black, M., & Hickey, R. (2008, July). Toward an understanding of flow in video games. *Computers in Entertainment, 6*, Article 20, 27 p. Retrieved from http://doi.acm.org/10.1145/1371216.1371223
Cowley, B., Heikura, T., & Ravaja, N. (2013). Learning loops—Interactions between guided reflection and experience-based learning in a serious game activity. *Journal of Computer Assisted Learning, 29*, 348–370.
Cowley, B., Ravaja, N., & Heikura, T. (2013). Cardiovascular physiology predicts learning effects in a serious game activity. *Computers & Education, 60*, 299–309. doi:10.1016/j.compedu.2012.07.014
Cowley, B., Kosunen, I., Lankoski, P., Kivikangas, J. M., Jarvela, S., Ekman, I., ... Ravaja, N. (2013). Experience assessment and design in the analysis of gameplay. *Simulation & Gaming, 45*, 41–69. doi:10.1177/1046878113513936
Cowley, B., Fantato, M., Jennett, C., Ruskov, M., & Ravaja, N. (2014). Learning when serious—Psychophysiological evaluation of a technology-enhanced learning game. *Journal of Educational Technology & Society, 17*, 3–16.
Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 159–181). Cambridge: Cambridge University Press.
Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Nordic Journal of Digital Literacy, 1*, 184–213.
Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed., p. 139). London: Sage.
Gable, P. A., & Harmon-Jones, E. (2008). Approach-motivated positive affect reduces breadth of attention. *Psychological Science, 19*, 476–482. http://dx.doi.org/10.1111/j.1467-9280.2008.02112.x
Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave Macmillan.
Gonzalez, C., & Czlonka, L. (2010). Games for peace: Empirical investigations with peacemaker. In J. Cannon-Bowers & C. Bowers (Eds.), *Serious game design and development: Technologies for training and learning* (pp. 134–149). Hershey, PA: IGI Global. doi:10.4018/978-1-61520-739-8
Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*, 468–484. http://dx.doi.org/10.1016/0013-4694(83)90135-9

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Gribble, J. D., Meyer, L., & Jones, A. (2003). *Quantifying and assessing learning objectives*. Melbourne. Retrieved from http://repository.unimelb.edu.au/10187/665

Groos, K. (1898). *The play of animals*. London: Chapman & Hall. http://dx.doi.org/10.1037/12894-000

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., & Schnitzler, A. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the National Academy of Sciences, 101*, 13050–13055. doi:10.1073/pnas.0404944101

Guillén-Nieto, V., & Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of it's a deal! *Computers & Education, 58*, 435–448. doi:10.1016/j.compedu.2011.07.015

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences., 20*, 169–206. http://dx.doi.org/10.1080/10508406.2010.508029

Harmon-Jones, E. (2003). Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology, 40*, 838–848. http://dx.doi.org/10.1111/psyp.2003.40.issue-6

Jann, K., Kottlow, M., Dierks, T., Boesch, C., & Koenig, T. (2010). Topographic electrophysiological signatures of fMRI resting state networks. *PLoS ONE, 5*, e12945. doi:10.1371/journal.pone.0012945

Kamzanova, A. (2011). EEG indices to time-on-task effects and to a workload manipulation (Cueing). *Proceedings of World Academy of Science, Engineering and Technology, 80*, 19–22.

Karakaş, S., Erzengin, Ö. U., & Başar, E. (2000). The genesis of human event-related responses explained through the theory of oscillatory neural assemblies. *Neuroscience Letters, 285*, 45–48. doi:10.1016/s0304-3940(00)01022-3

Kirschner, P., & Clark, R. (2006). Work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Kivikangas, J. M., Chanel, G., Cowley, B., Ekman, I., Salminen, M., Järvelä, S., & Ravaja, N. (2011). A review of the use of psychophysiological methods in game research. *Journal of Gaming and Virtual Worlds, 3*, 181–199. doi:10.1386/jgvw.3.3.181

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*, 169–195. http://dx.doi.org/10.1016/S0165-0173(98)00056-3

Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience and Biobehavioral Reviews, 36*, 677–695. doi:10.1016/j.neubiorev.2011.10.002

Koster, R. (2005). *A theory of fun for game design*. Scottsdale, AZ: Paraglyph Press.

Krause, C. M. (2000). Auditorily elicited event-related desynchronization (ERD) and synchronization (ERS) as a method of studying cortical correlates of cognitive processes. *Revista Espanola Neuropsicologia, 2*, 77–91.

Liang, K.-Y., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22. http://dx.doi.org/10.1093/biomet/73.1.13

Lord, T. R., French, D. P., & Crow, L. W. (2009). *College science teachers guide to assessment*. Arlington, VA: National Science Teachers Association.

Malone, T. W. (1981). What makes computer games fun? [for education]. *BYTE, 6*, 258–277.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies., 65*, 329–347. http://dx.doi.org/10.1016/j.ijhcs.2006.11.011

Markovska-Simoska, S., & Pop-Jordanova, N. (2010). Quantitative EEG characteristics of attention deficit hyperactivity disorder in adults. *Macedonian Journal of Medical Sciences, 3*, 368–377. doi:10.3889/MJMS.1857-5773.2010.0124

Mathewson, K. E., Basak, C., Maclin, E. L., Low, K. A., Boot, W. R., Kramer, A. F., … Fabiani, M. (2012). Different slopes for different folks: Alpha and delta EEG power predict subsequent video game learning rate and improvements in cognitive control tasks. *Psychophysiology, 49*, 1558–1570. doi:10.1111/j.1469-8986.2012.01474.x

McGilchrist, I. (2009). *The master and his emissary: The divided brain and the making of the Western world*. New Haven, CT: Yale University Press.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Greischar, L. L., & Davidson, R. J. (2009). Validation of regression-based myogenic correction techniques for scalp and source-localized EEG. *Psychophysiology, 46*, 578–592. doi:10.1111/j.1469-8986.2009.00787.x

Minkwitz, J., Trenner, M. U., Sander, C., Olbrich, S., Sheldrick, A. J., Schönknecht, P., … Hegerl, U. (2011). Prestimulus vigilance predicts response speed in an easy visual discrimination task. *Behavioral and Brain Functions, 7*, 1–8. doi:10.1186/1744-9081-7-31

Niedermeyer, E. (2005). *Electroencephalography: Basic principles, clinical applications, and related fields* (5th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *Curriculum Journal, 16*, 455–474. doi:10.1080/09585170500384529

Palomäki, J., Kivikangas, M., Alafuzoff, A., Hakala, T., & Krause, C. M. (2012). Brain oscillatory 4–35 Hz EEG responses during an n-back task with complex visual stimuli. *Neuroscience Letters, 516*, 141–145. doi:10.1016/j.neulet.2012.03.076

Panksepp, J. (1998). *Affective neuroscience : The foundations of human and animal emotions*. New York, NY: Oxford University Press.

Pizzagalli, D. A., Sherwood, R. J., Henriques, J. B., & Davidson, R. J. (2005). Frontal brain asymmetry and reward responsiveness: A source-localization study. *Psychological Science, 16*, 805–813. doi:10.1111/j.1467-9280.2005.01618.x

Pope, A. T., & Palsson, O. S. (2001). *Helping video games "rewire our minds"*. Hampton, VA: NASA Research. Retrieved from http://archive.org/details/nasa_techdoc_20040086464

Ravaja, N., Saari, T., Salminen, M., Laarni, J., & Kallinen, K. (2006). Phasic emotional reactions to video game events: A psychophysiological investigation. *Media Psychology, 8*, 343–367. http://dx.doi.org/10.1207/s1532785xmep0804_2

Rotenberg, V. S., & Arshavsky, V. V. (1997). Right and left brain hemispheres activation in the representatives of two different cultures. *Homeostasis, 38*, 49–57.

Roth, B. (1961). The clinical and theoretical importance of EEG rhythms corresponding to states of lowered vigilance. *Electroencephalography and Clinical Neurophysiology, 13*, 395–399. http://dx.doi.org/10.1016/0013-4694(61)90008-6

Sage, G. H., & Bennett, B. (1973). The effects of induced arousal on learning and performance of a pursuit motor skill. *Research Quarterly. American Association for Health, Physical Education and Recreation, 44*, 140–149. doi:10.1080/10671188.1973.10615189

Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience, 12*, 1370–1371. doi:10.1038/nn.2412

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher, 35*, 19–29.10.3102/0013189X035008019

Staude-Müller, F., Bliesener, T., & Luthman, S. (2008). Hostile and hardened? An experimental study on (de-) sensitization to violence and suffering through playing video games. *Swiss Journal of Psychology, 67*, 41–50. doi:10.1024/1421-0185.67.1.41

Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education, 63*, 43–51. doi:10.1016/j.compedu.2012.11.019
Sutton-Smith, B. (1997). *The ambiguity of play*. Cambridge, MA: Harvard University Press.
Wamsley, E. J., Tucker, M. A., Payne, J. D., & Stickgold, R. (2010). A brief nap is beneficial for human route-learning: The role of navigation experience and EEG spectral power. *Learning & Memory, 17*, 332–336.
Wang, Q., Sourina, O., & Nguyen, M. (2011). Fractal dimension based neurofeedback in serious games. *The Visual Computer, 27*, 299–309. doi:10.1007/s00371-011-0551-5
van Oyen Witvliet, C., & Vrana, S. R. (1995). Psychophysiological responses as indices of affective dimensions. *Psychophysiology, 32*, 436–443. http://dx.doi.org/10.1111/psyp.1995.32.issue-5

Appendix A

Table A1 outlines the relationship between types of questions, the Bloom level assigned to them and the game data or experience which the question addresses—it also lists the number of such questions asked. A sample list of questions and complete assessment protocols follow.

Table A1. Relationship between types of question, their Bloom level, and the game

Type of question	Bloom level	Game data covered	Number of question
Recall/multiple choice questions	1 or 2	Geography, groups and leaders, polls, tutorial, timeline, intro videos	28
Application/active recall	3	The effect of actions available to the player:	5
		Security: retaliation, suppression, public order, prisoners	
		Politics: international, domestic, cabinet, Palestinian Authority, workers, trade	
		Construction: aid (medical, infrastructure, civil), domestic, settlements, wall, refugees	
Analysis of given data, requiring inference or induction	4	Sequences of actions where the outcome can be defined with certainty, given the particular state/stage of the game. Which actions do or don't work together	4 (28)
Evaluation or judgement of given game scenarios	5	Strategies in the game with their probable outcomes	4
Synthesis of game data, or use of game scenario, to relate real world situation	6	Personal expressions/descriptions of topical knowledge/lessons from game experience	6

Notes: Final column is the number of questions of that level that the player had to answer (the number in parentheses shows how many separate items were asked in multi-part rating questions).

The following is a representative sample list of questions used to assess player learning, numbered by the order they appeared to participants. They are labelled with their Bloom level: see Table A1. Following the questions below, we give details of the assessment protocols for each question type.

Qualitative questions

These open-answer questions are given to assess the higher levels of subject learning.

1 What is your current understanding of the causes for the Israeli–Palestine conflict?

— BLOOM 6

4 Give your current estimate of the possibility of future peace in the region. Indicate the probability using a number from 1 to 9, with 1 being impossible and 9 being absolutely certain

— BLOOM 6

Downloaded by [213.27.242.194] at 02:00 06 October 2014

	Highly improbable	2	3	4	Unknowable	6	7	8	Highly probable
Peace									

(i) Describe why you picked that number.

Quantitative questions

Below, we list a sample of the quantitative questions. The answer to the question is listed directly below it. Following that is the assumptions behind the question—these include any assumption supporting the validity of the answer, plus the necessary condition for the question to work in the experiment i.e. how the player learns the information.

10	"A high level of trust and mutual respect between the Israeli PM and Palestinian Authority President is vital to successfully build peace."	
	Rate the following actions/policies by the Israeli PM for their effectiveness in building trust.	
	a) Security: securing/patrolling areas with police	1 2 3 4 5
	b) Politics: political speech-making	1 2 3 4 5
	c) Construction: expand Israeli settlements	1 2 3 4 5
	d) Security: engage and destroy militant targets	1 2 3 4 5
	e) Construction: authorise aid packages to the Palestinians	1 2 3 4 5
	f) Politics: push for improved commercial and cultural relations	1 2 3 4 5
	g) Security: restrict and curtail Palestinian population	1 2 3 4 5
	h) Politics: negotiation and consultation between leaders	1 2 3 4 5
	i) Construction: Israeli domestic welfare initiatives	1 2 3 4 5
(answer)	Ratings given in two answers (x, y) are evaluated by $(y-x)\times w$, where w is given below:	
	$a=-1$, $b=1$, $c=-1$, $d=-1$, $e=1$, $f=1$, $g=-1$, $h=1$, $i=-1$	
	These weights were derived from the responses of the AI to actions corresponding to those named, in the games played by test participants	
	We have estimated as follows:	
	$a=3$, $b=4$, $c=1$, $d=2$, $e=3$, $f=5$, $g=1$, $h=4$, $i=2$	
(assumptions)	We assume the correctness of the answer based on observation/play. Player can infer from observing relevant variables while trying this strategy—BLOOM 4	

11	Which of the following regional countries share a border with the state of Israel?	
	Cyprus	[no]
	Egypt	[yes]
	Iraq	[no]
	Jordan	[yes]
	Lebanon	[yes]
	Saudi Arabia	[no]
	Syria	[yes]
	Turkey	[no]
	Yemen	[no]
(answer)	Egypt, Jordan, Lebanon, Syria	
(assumptions)	Player observes large-scale map in game, does not know relevant geography—BLOOM 1	

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Downloaded by [213.27.242.194] at 02:00 06 October 2014

13	Name two fears of the Israeli public
(*answer*)	*Palestinian militant attacks,*
	instability,
	economic loss
(assumptions)	Player reads about group—BLOOM 2
14	Below are four examples of a sequence of actions/policies that can be done by the Israeli Prime Minister. Tick each sequence if you believe that the final effect (of all the actions in it) would be to please BOTH the Israeli and Palestinian sides and overcome the zero-sum effect
	In the Israeli–Palestine conflict, as in the game, it is often the case that a particular action or policy by a leader will be disapproved by one side as much as it is approved by the other. This is known as the zero-sum effect
	Now rate each sequence for how well it would please both sides at the same time
	The score indicates how pleased both sides are after all the actions in the sequence are done. So a score of 1 counts as "really displeases one or both sides" and score of 5 counts as "really pleases both sides'
	a) 1 2 3 4 5
	Increase the number of Police Patrols
	Order the Israeli Army to destroy militant infrastructure
	Engage in political talks with the Palestinian President to demand anti-militant action
	Give direct funding to construct Palestinian medical institutions
	b) 1 2 3 4 5
	Order the removal of existing Jewish settlements
	Enact security policy to reduce Palestinian curfews
	Enact policy to decrease trade restrictions
	Enact a domestic economic stimulus package
	c) 1 2 3 4 5
	Engage in political talks with Palestinian President to ask for anti-militant support
	Increase the number of Police Patrols
	Enact policy to decrease trade restrictions
	Order the Israeli Army to secure an area
	d) 1 2 3 4 5
	Enact a policy to allow Palestinian refugees to immigrate back to their homeland
	Enact an initiative for more cross-cultural projects
	Through the UN, give funding for Palestinian Medical institutions
	Make a speech to the Israeli people calling for cooperation with Palestinians
(answer)	Where scores are written (*Israeli Approval*, *Palestinian Approval*), test-playing the above action sequences (from the first move of the game) achieved the following scores:
	$a = (15, -25)$, $b = (-6, 23)$, $c = (8, 5)$, $d = (-5, 5)/(-4, 5)/(-3, 10)$
	Rating each action individually across all played games, and averaging, gave scores of:
	$a = (4.69, -3.81)$, $b = (-0, 8, 8.56)$, $c = (3.77, 1.74)$, $d = (-.29, 3.25)$
	Only c) should be ticked, and the ratings given (x, y) are evaluated by $(x - y) \times w$, where w is given below:
	$a = -1$, $b = -1$, $c = 1$, $d = -1$

Downloaded by [213.27.242.194] at 02:00 06 October 2014

(assumptions)	We assume the correctness of the answer based on observation/play. Each strategy was tested three times. Player can infer from observing relevant variables while playing—BLOOM 5
19	Of all the interested parties (represented in the game as groups and leaders), [_____] are most opposed to your plans (i.e. have the lowest approval of you in the game)
(answer)	Militants—Name of any one should suffice e.g. Hamas
(assumptions)	Player notes the approval ratings at beginning of game—BLOOM 2
32	What is the general feeling among the Israeli public towards the Palestinian public (as simulated at the beginning of the game)?
	a) Apathy b) Anger c) Despair d) Compassion
(answer)	a)
(assumptions)	Player sees the event reporting this info, or infers from Israeli sympathy poll—BLOOM 1
33	What could be done to improve this? Tick any that apply
	a) Focus on securing of borders and segregation
	b) Focus on diplomatic relations with foreign countries
	c) Focus on intercultural initiatives and trade
	d) It is impossible
	e) Focus on Israeli settler issues
	f) Focus on domestic education and welfare
	g) Focus on destruction of militants
	h) Focus on green issues
(answer)	a) and c), and to a lesser degree f)
(assumptions)	Player infers from Israeli sympathy poll after playing correct strategy—BLOOM 3
	This question should only be scored if Q32 was answered correctly

Assessment protocol

Note: gain scores are potentially negative—if answers go from right to wrong, they are given negative points. However, this "negative learning" score can be treated as zero in post-processing, achieving the same effect as an initial assumption of no negative learning in an exploratory analysis.

For questions (of level 1–5) that requested specific information but allowed open answers (free text input), we defined a synonymy set, that is, a set of answers which could legitimately be given in lieu of the "correct" answers.

Rating questions were assessed by a formula (explained below) that preserved the magnitude of the subject's response preference without giving an arbitrary "truth" value to the rating item.

All level 1–5 questions thus obtained a gain score. These were then weighted. Initially, weights were the product of the gain score and the number of the Bloom level, which gives a linear increase in importance over Bloom levels. Yet the "learning value" of the Bloom levels is not defined in a scalar sense, only as ordinals, so there is more than one option supported by theory for weighting each level. For instance, the importance of learning at higher levels could be considered parametrically greater than lower levels (because mastery at each level is considered to require mastery at all the lower levels first): applying this changes the weight values from linear scaling [1, 2, 3, 4, 5, 6] to exponential scaling [1, 2, 4, 8, 16, 32].

Downloaded by [213.27.242.194] at 02:00 06 October 2014

- For single response multiple choice (e.g. Q32).
  - 1st and 2nd response are the same = 0 points.
  - 2nd response is correct and 1st response is not = 1 point.
  - 1st response is correct and 2nd response is not = −1 point.
- For multi-response multiple choice (e.g. Q11).
  - For every response that is the same both times = 0 points.
  - For every correct response in 2nd answer (that is not in 1st answer) = 1 point.
  - Every correct response in 1st answer (that is not in 2nd answer) = −1 point.
  - Every incorrect response in 1st answer (that is not in 2nd answer) = 1 point.
  - Every incorrect response in 2nd answer (that is not in 1st answer) = −1 point.
- For single answer "open" questions (e.g. Q19)—the right answer, or a synonym, *or a competing but equally correct answer*,[3] is in 2nd response but not in 1st response = 1 point.
- For multi-answer "open" questions (e.g. Q13)—every correct answer, or a synonym, *or a competing but equally correct answer*[1], in 2nd response that is not in 1st response = 1 point.
- Rating questions (e.g. Q10) are assessed by an objective formula: $(y-x)\times w$,

where x, y and w are defined as follows (refer also to question 10 above).

  - So, for example, in this one rating-type question we have these nine items, with a weight attached (either −1, 0, 1) which was derived from the data of game players by asking, for each rating item, what was the reaction in the variable of interest after the action that is cited in the rating item (in question 10, the variable of interest is the relationship between Israeli and Palestinian leaders, defined by a scalar in the game).

    Thus, we do not pre-judge what score the rating should be, but rather only whether the action associated with the rating was positive, negative or neutral (with respect to the question asked). This is defined by our weights w.

By subtracting the first score from the second, we get a magnitude and a sign. Say in item 10.a (with weight −1), the subject responds first with 4, second with 2. Then, the calculation would be

$$(2-4)\times -1 = 2$$

The subject has downgraded his rating of that action (which was defined as a bad action for the purpose of building trust, based on the data), from more positive (4) to more negative (2), so his score is +2, preserving the magnitude of the change. If he had answered in the opposite way, first 2 and second 4, he would be upgrading his estimate of the quality of the (bad) action, and thus would get a score of

$$(4-2)\times -1 = -2$$

Thus, we preserve magnitude without giving an ad hoc "true" value to the rating item.

- The procedure for assessing open questions is detailed in the next section.

Open question assessment

From the 41 questions, 6 were open questions of the form: "What is your understanding of [a topic]?" or "Describe why you [responded to the antecedent quantitative question as they did]?" These open questions were analysed separately, since they were not held to be immediately comparable to the quantitative questions in terms of scoring. They represented opportunity for wider contemplation when answering and thus enabled responses that might (or might not) be evaluated as containing Bloom's "level 6" knowledge.

Downloaded by [213.27.242.194] at 02:00 06 October 2014

Supplying open questions and evaluating them separately was aimed at investigating whether being asked to reflect on the game play would improve and deepen understanding of the topic, especially in a social group where the "social norm" would help prevent the participants from reflecting off-topic (day-dreaming). The evidence was expected to present in the difference between the quality of first- and second-round responses to the six open questions.

This difference was made visible by applying evidential criteria to assessing the match between pre- and post-game answers and "level 6" of Bloom's taxonomy. In other words, each of the 6 criteria below was applied to the material at hand, once for the first answer and once for the second, and the analyses were compared.

(1) Indications of finding central factors or components of a phenomenon, understanding how they are related to one another, can compare phenomenon to other similar phenomena, ponders upon the differences, commonalities and reasons for differences.

(2) Indications of being able to take into account different perspectives (Israel, Palestine, rest of the world).

(3) Indications of being able to approach the phenomena from different themes (economy, politics, religion, power, history, culture ...).

(4) Indications of being able to apply a principle derived from or presented in the game to another context or indications of being able to import principles (e.g. to pieces of one's experience and knowledge domain) from one's other contexts into the game thematic context.

(5) Indications of being able to reflect upon the principles presented in the game and assessing their validity and applicability in other contexts, evaluating the boundaries and terms of the applicability e.g. in relation to time and place.

(6) Indications of self-reflection or questioning of one's own beliefs.

All responses to 6 open questions of 45 participants were first filtered according to the simple constraint of meeting the Bloom level 6 description cited above. As predicted, only a small percentage of answers per question succeeded to meet the criteria in their content; most of the responses consisted of few sentences without indications of wider contemplation of the topic.

Those qualified responses were then evaluated by manual assessment, according to the six detailed criteria above, maintaining consistency and impartiality by using two separately working assessors following a fixed protocol for each question. After this parallel reflection phase, a concerted discussion was held to evaluate how well and in what proportion did the qualified responses meet the criteria, which also determined the inter-rater reliability in addressing the criteria to the responses.

© 2014 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 3.0 license.
You are free to:
Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material for any purpose, even commercially.
The licensor cannot revoke these freedoms as long as you follow the license terms.
Under the following terms:
Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
No additional restrictions
You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.