ABSTRACT

The field of games-based learning (GBL) has a dearth of empirical evidence supporting the validity of the approach (Connolly, Stansfield, & Hainey, 2007a; de Freitas, 2006). One primary reason for this is a distinct lack of frameworks for GBL evaluation. The literature has a wealth of articles suggesting ways that GBL can be evaluated against particular criteria with various experimental designs and analytical techniques. Based on a review of existing frameworks applicable to GBL and an extensive literature search to identify measurements that have been taken in relevant studies, this chapter will provide general guidelines to focus researchers on particular categories of evaluation, individual measurements, experimental designs and texts in the literature that have some form of empirical evidence or framework relevant to researchers evaluating GBL environments particularly focusing on learner performance. A new evaluation framework will be presented based on the compilation of all the particular areas and analytical measurements found in the literature.
INTRODUCTION

One of the primary concerns associated with the GBL literature is the dearth of empirical evidence supporting the validity of the approach (Connolly, Stansfield, & Hainey, 2007a; de Freitas, 2006). O’Neil et al (2005) believe that an essential element missing is the ability to properly evaluate games for education and training purposes. If games are not properly evaluated and concrete empirical evidence is not obtained in individual learning scenarios that can produce generalisable results, then the potential of games in learning can always be dismissed as unsubstantiated optimism.

In the O’Neil study, a large amount of literature was collected and analysed from the PsycINFO, EducationAbs, and SocialAciAbs information systems. Out of several thousand articles, 19 met the specified criteria for inclusion and had some kind of empirical information that was either qualitative, quantitative or both. The literature was then viewed through Kirkpatrick’s four levels for evaluating training and the augmented CRESST model. The majority of the studies reviewed analysed performance on game measurements. Other studies included observation of military tactics used, observation, time to complete the game, transfer test of position location, flight performance and a variety of questionnaires including exit, stress and motivation questionnaires.

The review of empirical evidence on the benefit of games and simulations for educational purposes is a recurring theme in the literature and can be traced even further back. For example, Randel, Morris, Wetzel, and Whitehill (1992) examined 68 studies from 1963 comparing simulations/games approaches and conventional instruction in direct relation to student performance. Some of the following main discoveries were made:

- 38 (56%) of the studies found no difference; 22 (32%) of the studies found a difference that favoured simulations/games; 5 (7%) of studies favoured simulations/games however control was questionable; 3 (5%) found differences that favoured conventional instruction.
- With regards to retention simulations/games induced greater retention over time than conventional techniques.
- With regards to interest, out of 14 (21%) studies, 12 (86%) showed a greater interest in games and simulations over conventional approaches.

Although lack of empirical evidence supporting GBL is not a new issue, the growing popularity of computer games in conjunction with recent advances in games and hardware technology, the emergence of virtual worlds and massively multiplayer online games (MMOGs), reinforces the need for a flexible evaluation framework that can be used by Evaluation researchers. This chapter presents such an evaluation framework.

In the next section, we examine previous research and, in particular, discuss the types of evaluation that can be used and their applicability and importance in the field of GBL. We also examine previous evaluation frameworks that have been presented in the literature that could be applicable to GBL and follow that with the results of an extensive literature review to identify studies that performed some form of evaluation and attempted to take appropriate measurements through various experimental designs particularly focusing on learner performance. On the basis of these reviews, we then present a flexible framework for GBL evaluation from a pedagogical perspective. In the final section we discuss future validation of the new framework.

PREVIOUS RESEARCH

Evaluation

Ainsworth (2003) divides evaluation into two main types: evaluation to inform design encompassing