

Evaluating online learning: a review of studies on e-testing in the university context

Valutare l'apprendimento online: una rassegna degli studi sull'e-testing nel contesto universitario

Andrea Nardi^a

^a *Università degli Studi di Firenze*, andrea.nardi@unifi.it

Abstract

Nowadays, the class size and the growing number of undergraduate students generate unprecedented problems of management, monitoring and evaluation of learning. From this point of view technologies may provide opportunities which deserve to be investigated for their potential to contribute to solve problems related to the management of large size classes. Since many teachers show resistance towards the replacement of traditional exams with e-assessment, little effort has been made for the innovation of the evaluation procedures at the university level. However, several studies have been carried out in this field. This paper provides an overview of the current electronic evaluation activity in the university context. Data shows benefits related to students' monitoring for the simplification of procedures and the reduction of teachers' cognitive load. Moreover students appreciate the electronic mode for the possibility to obtain immediate feedback. The challenges concern the technological infrastructure, the scalability and safe use.

Keywords: online testing; online exam; online assessment; university.

Abstract

Oggi giorno le dimensioni delle classi e il numero crescente di studenti in ambito universitario pongono inediti problemi di gestione, monitoraggio e valutazione degli apprendimenti. La tecnologia da questo punto di vista rappresenta un'opportunità che necessita di essere indagata perché potenzialmente in grado di contribuire alla risoluzione delle problematiche connesse alla gestione di una grande mole di iscritti. Poiché molti docenti manifestano diffidenza verso la sostituzione degli esami tradizionali con modalità elettroniche di valutazione, pochi sforzi sono stati compiuti in questi anni per l'innovazione delle procedure valutative. Tuttavia, molti sono gli studi realizzati in questo campo. Questo lavoro fornisce una panoramica dell'attuale attività di valutazione elettronica nel contesto universitario. Dai dati raccolti emergono benefici in relazione al monitoraggio degli studenti per la semplificazione delle procedure e la riduzione del carico di lavoro dei docenti. Inoltre gli studenti apprezzano la modalità elettronica per la possibilità di ottenere feedback immediati. Le sfide riguardano l'infrastruttura tecnologica, la scalabilità e la sicurezza.

Parole chiave: test online; esami online; valutazione online; università.

1. Introduzione

La storia dei test basati su computer inizia nei primi anni Settanta quando con l'introduzione dei primi calcolatori elettronici in ambito accademico s'iniziano a esplorare le potenzialità dell'uso della tecnologia nella progettazione di nuovi ambienti di apprendimento e test valutativi. Dapprima sviluppati in ambito militare, i test computerizzati iniziano ad essere applicati per la valutazione e la pratica degli esami negli anni Novanta quando vengono implementati i primi programmi di test su larga scala per l'abilitazione professionale, la certificazione, le ammissioni e i test psicologici. Inizialmente le prove vengono svolte principalmente mediante l'uso di programmi specifici e forme primitive di test computerizzati di tipo adattivo (CAT), mentre per i primi test online si deve attendere l'inizio del 2000 (Luecht & Sireci, 2011).

Diverse sono le espressioni che rimandano, in modo più o meno specifico, all'uso degli strumenti digitali per la valutazione. Per riferirsi genericamente all'uso delle tecnologie ICT nel processo valutativo si utilizza il termine E-Assessment (Rout & Patnaik, 2011). L'espressione Computer-Based Assessment (CBA) viene utilizzata più specificatamente per la valutazione automatica delle risposte fornite dagli studenti. Il Computer-Based Testing (CBT) consiste nella somministrazione di test elettronici mediante un computer o un dispositivo equivalente anche in modalità offline, mentre l'Online Testing o Web-based testing (WBT) prevede la somministrazione di prove attraverso internet. In ambito universitario si utilizza il termine esame elettronico (e-Exam) per riferirsi a una valutazione sommativa controllata a tempo, condotta utilizzando il computer di ciascun candidato, fornita attraverso un sistema dedicato, inclusa come parte di un modulo LMS (Kuikka, Kitola & Laakso, 2014) o svolta in modalità off-site attraverso l'utilizzo di un'unità flash USB (Fluck & Hillier, 2016).

Guardando allo sviluppo storico è possibile rintracciare almeno quattro generazioni di e-assessment o Computer-Based Assessment (CBA) (Redecker & Johannessen, 2013):

- *test computerizzati*. Si tratta in genere di test a risposta multipla o esercizi chiusi gestiti tramite computer che, rispetto alle prove tradizionali carta e matita, consentono l'automazione di alcuni processi legati alla valutazione (ad es. consegna, correzione, assegnazione dei punteggi);
- *test adattivi computerizzati*. Il sistema adatta la difficoltà e il timing degli item in base alle risposte degli esaminandi, selezionando gli elementi del test e adeguandoli al livello di conoscenza e abilità dello studente. I quesiti vengono generalmente prelevati da un database e grazie a strumenti statistici più o meno sofisticati, il sistema assegna i diversi compiti;
- *misurazione continua*. Il sistema utilizza misure calibrate in modo continuo e non intrusivo per stimare i cambiamenti nel rendimento dello studente e proporre forme di valutazione più personalizzate. Grazie all'analisi dinamica e all'interpretazione automatica dei dati ottenuti, gli studenti sono monitorati e guidati dall'ambiente elettronico che usano per le loro attività di apprendimento.
- *misurazione intelligente*. Il sistema genera punteggi intelligenti, interpreta i profili individuali e fornisce consigli a studenti e docenti mediante basi di conoscenza e procedure di inferenza. Tecniche di data-mining possono essere utilizzate per il monitoraggio e il tutoraggio individuale dello studente (un sistema può ad esempio identificare gli studenti a rischio di drop out e generare report diagnostici).

Oggi l'uso di strumenti elettronici a supporto della valutazione sembra ricevere una rinnovata attenzione. Il fenomeno delle classi numerose, infatti, che contraddistinguono molti

degli attuali contesti universitari nazionali ed internazionali, pone inedite esigenze valutative. La *massificazione* dell'istruzione universitaria, il crescente numero di esaminandi, l'uso sempre più frequente dei dispositivi mobili per l'apprendimento, lo spostamento di parte dell'attività didattica e gestionale all'interno di ambienti online e classi virtuali, richiedono la riformulazione dei tradizionali strumenti e metodi di monitoraggio degli apprendimenti, così come ulteriori canali di valutazione e forme più *snelle* di assegnazione dei punteggi. Non è un caso quindi che un numero crescente di istituti di istruzione superiore e di università abbia iniziato ad utilizzare strumenti elettronici e web-based per la valutazione dei propri studenti (Kuikka et al., 2014).

D'altra parte, i benefici della somministrazione di esami in modalità digitale sono noti da tempo a livello internazionale e molti test standardizzati e valutazioni con un'alta posta in gioco (*high-stakes assessments*) vengono oggi effettuati principalmente al computer e online. Mentre a livello internazionale abbiamo già approcci consolidati per quanto riguarda la secondaria superiore, come nel caso dell'Organization for Economic Co-operation and Development (OECD), a livello universitario mancano modelli condivisi e pratiche diffuse. In Italia le prove al computer sono già utilizzate da molti Centri Linguistici di Ateneo (CLA) per la certificazione linguistica, per gli esami di accertamento delle competenze informatiche, ma sembrano ancora poco diffuse in ambito universitario.

L'introduzione degli esami elettronici pone indubbiamente sfide inedite e trova spesso ancora resistenze da parte dei docenti (Betlej, 2013; Kuikka et al., 2014) in ragione di tre principali fattori (Sindre & Vegendla, 2015):

- *infrastruttura tecnologica*. Le università devono affrontare significativi costi di start-up, tra cui hardware, software, connettività, formazione del personale, supporto tecnico e sviluppo di adeguati livelli di alfabetizzazione digitale;
- *scalabilità*. Svolgere le prova in modalità computerizzata è particolarmente oneroso in termini di risorse, specialmente se i test vengono svolti su attrezzature fornite dall'università;
- *sicurezza*. Vanno previste misure per contrastare i fenomeni di e-cheating, individuare i plagi, sorvegliare gli studenti, gestire le criticità relative ai processi di autenticazione dello studente, proteggere l'identità e i dati sensibili, limitare i problemi tecnici e di connettività, gestire le difficoltà relative all'interoperabilità, alla riusabilità, alle procedure di aggiornamento, supporto e manutenzione di software e hardware.

Ci sono tuttavia alcuni benefici della valutazione elettronica che possono rappresentare un'occasione di rinnovamento delle tradizionali pratiche valutative e che la rendono una risorsa particolarmente utile nel caso di classi numerose (Dermo, 2009; JISC, 2010; Redecker & Johannessen, 2013; Sindre & Vegendla, 2015):

- *sviluppo dei test*. Nuove strategie per far fronte a grandi numeri di candidati; risparmio di tempo e risorse; generazione automatica delle domande; condivisione di domande tra istituti attraverso database; possibilità di sviluppare una gamma più ampia di item, prove e compiti più complessi (processi cognitivi di alto ordine e competenze); domande adattive e possibilità di venire incontro a livelli diversi di abilità individuali come nel caso di studenti con disabilità o con diverso background linguistico; elementi multimediali e simulazioni che permettono di misurare competenze non facilmente valutabili dai test tradizionali;
- *consegna dell'esame*. Riduzione dei costi di distribuzione, somministrazione e raccolta dei test; standardizzazione delle procedure di prova; feedback mirati e

istantanei; analisi automatica dei dati degli studenti (andamento statistico, progressi dell'intera classe, grafici di sintesi); sistemi di profilazione degli studenti; possibilità di sostenere le prove a distanza; aumento della sicurezza dovuto alla trasmissione elettronica e alla criptazione (i test possono essere inviati via Internet all'ultimo minuto, riducendo la possibilità di esporre gli studenti alle domande prima della prova); correzione di eventuali errori nei test in corsa;

- *valutazione*. Riduzione dei costi relativi alla distribuzione delle risposte; automazione nella assegnazione dei voti; maggiore precisione nell'assegnazione dei punteggi; processo di revisione dei test più rapido e maggiormente controllato; riduzione degli errori sui libretti; eliminazione delle problematiche derivanti da materiale perso o danneggiato; maggiore leggibilità delle risposte fornite dai candidati e migliore protezione della loro identità.

L'obiettivo del presente lavoro è quello di esplorare le potenzialità e i limiti degli esami elettronici (CBT) con un focus sugli aspetti relativi all'online (WBT) rispetto alla tradizionale modalità cartacea (PBT) allo scopo di verificare se, e fino a che punto, la modalità elettronica possa nel tempo diventare un'effettiva alternativa al cartaceo, permettendo di gestire il processo valutativo in modo più efficiente, specie nel contesto delle classi numerose in contesto accademico.

2. Strategia di ricerca

Il lavoro ha previsto un'analisi della letteratura, con particolare attenzione alle ricerche più recenti svolte in ambito universitario e quindi in grado di informare maggiormente sugli aspetti relativi all'attuale contesto tecnologico e sociale nel quale la valutazione elettronica si inserisce. La ricerca si è avvalsa perlopiù di Google Scholar, un motore di ricerca basato su un sistema di indicizzazione in grado di restituire un ampio numero di risultati contenuti in database specializzati (ad esempio ERIC, Web of Science) o in riviste internazionali accreditate (ad esempio Taylor & Francis, Wiley, Sage, Elsevier). La consultazione e la selezione delle risorse hanno seguito un approccio metodologico ispirato all'Evidence Based Education nella sua declinazione *informed* o *aware* (Calvani, 2013). Per interrogare i motori di ricerca sono state utilizzate specifiche query volte a circoscrivere l'ambito. In prima battuta sono state utilizzate espressioni in lingua italiana come «valutazione online», «esami online» «test online» ed espressioni di significato affine come «valutazione elettronica», «valutazione basata sul computer». Non essendo riusciti a trovare un numero consistente di risorse attinenti, si è deciso di ripetere la ricerca formulando le stesse query in lingua inglese: «online assessment», «online exam», «online testing» ed espressioni di significato affine come «web-based testing», «web-based exam», «web-based assessment», «computer-based assessment», «computer-based exam», «computer-based testing», «mobile assessment», «mobile exam», «mobile testing». L'uso di parole chiave in inglese ha prodotto migliori risultati. Gli studi sono stati inclusi sulla base dei seguenti criteri:

- articoli scientifici pubblicati su riviste specializzate, scritti in lingua italiana o inglese e tesi di dottorato pubblicati tra il 2000 e il 2018;
- meta-analisi, systematic review e critical review o, in ultima analisi, ricerche sperimentali o quasi sperimentali;
- studi condotti nel contesto dell'Alta formazione con studenti universitari;
- studi che prevedano una comparazione dei risultati e l'utilizzo di strumenti di valutazione computer-based o web-based in almeno una delle due condizioni;

- studi con una dichiarazione esplicita del focus e dei risultati perseguiti.

N°	Anno	Tema	Condizione
4	2000	Performance degli studenti (4)	CBT* vs PBT
6	2001	Performance degli studenti (5) Percezioni degli studenti (1)	CBT vs PBT
9	2002	Performance degli studenti (7) Percezioni degli studenti (2)	CBT vs PBT WBT vs PBT
5	2003	Performance degli studenti (4) Design e comparabilità dei test (1)	CBT vs PBT WBT vs PBT
10	2004	Performance degli studenti (8) Percezioni degli studenti (2)	CBT vs PBT WBT vs PBT
6	2005	Performance degli studenti (3) Percezioni degli studenti (1) Design e comparabilità dei test (1) Rassegna della letteratura (1)	CBT vs PBT WBT vs PBT
10	2006	Performance degli studenti (8) Percezioni degli studenti (1) Design e comparabilità dei test (1)	CBT vs PBT WBT vs PBT
4	2007	Performance degli studenti (3) Percezioni degli studenti (1)	CBT vs PBT WBT vs PBT
6	2008	Performance degli studenti (2) Percezioni degli studenti (2) Rassegna della letteratura (2)	CBT vs PBT WBT vs PBT
7	2009	Performance degli studenti (3) Percezioni degli studenti (2) Criticità riscontrate dagli studenti (2)	CBT vs PBT WBT vs PBT
6	2010	Performance degli studenti (5) Percezioni degli studenti (1)	CBT vs PBT WBT vs PBT
7	2011	Performance degli studenti (3) Percezioni degli studenti (2) Criticità riscontrate dagli studenti (2)	CBT vs PBT WBT vs PBT
6	2012	Performance degli studenti (2) Percezioni degli studenti (3) Criticità riscontrate dagli studenti (1)	CBT vs PBT WBT vs PBT
5	2013	Performance degli studenti (2) Criticità riscontrate dagli studenti (3)	CBT vs PBT WBT vs PBT
5	2014	Performance degli studenti (1) Percezioni degli studenti (3) BYOD (1)	CBT vs PBT WBT vs PBT
3	2015	Percezioni degli studenti (1) Rassegna della letteratura (1) BYOD (1)	WBT vs PBT
4	2016	Performance degli studenti (1) Criticità riscontrate dagli studenti (2) Design e comparabilità dei test (1)	WBT vs PBT
2	2017	Performance degli studenti (1) Percezioni degli studenti (1)	WBT vs PBT
Totale Studi 105	Periodo 2000-2017	Temi indagati 5	Condizioni 3

Figura 1. Studi che hanno comparato gli effetti di CBT (Computer-Based Testing), WBT (Web-Based-Testing) e PBT (Paper-Based Testing) e che rispettano i criteri di selezione.

Sono stati sottoposti a scansione oltre 3000 rapporti in base ai titoli e gli abstract, dei quali sono stati controllati quasi 500 studi full-text. Infine, sono risultati ammissibili, perché rispondevano ai criteri, 105 contributi. Nella Figura 1 sono riportate le principali informazioni relative alla lista completa degli articoli che hanno soddisfatto i criteri di

ricerca. In questa sede per limiti di spazio privilegeremo il resoconto dei risultati più significativi.

3. Analisi dei risultati

La valutazione computerizzata degli apprendimenti è stata oggetto di un crescente interesse negli ultimi 10 anni e la ricerca che ha analizzato gli effetti della somministrazione di prove al computer e online si è concentrata in particolare sui seguenti elementi di indagine: (i) la comparazione delle prestazioni misurate attraverso il confronto tra modalità tradizionali e nuove modalità di testing; (ii) l'atteggiamento degli studenti nei confronti dell'e-assessment; (iii) i fattori di criticità identificati dagli esaminandi; (iv) la progettazione, amministrazione e comparabilità dei test; (v) i recenti sviluppi in chiave BYOD (Bring-Your-Own Device).

3.1. Performance degli studenti

La ricerca che ha analizzato gli effetti dell'introduzione dei sistemi di valutazione elettronica si è concentrata fino ad ora soprattutto sull'analisi dell'impatto prodotto dal computer/web-based testing e dal paper-testing in termini di rendimento degli studenti (Walker & Handley, 2016). La rassegna della letteratura svolta da Noyes e Garland (2008) ha mostrato come con gli sviluppi della tecnologia, di più sofisticate misure comparative e di atteggiamenti più positivi da parte degli utenti, si stia ottenendo una situazione di sempre maggiore equiparabilità di performance tra le due tipologie di test (soprattutto per i compiti standardizzati come quelli a domande a scelta multipla chiusa). Rassegne di studi più recenti svolte a livello universitario hanno riportato tuttavia anche dati misti e discordanti (Boevé, Meijer, Albers, Beetsma & Bosker, 2015): in alcuni casi non vi sono significative differenze nei punteggi ottenuti dagli studenti (Jeong, 2014), in altri casi le performance risultano migliori nella condizione digitale (Nikou & Economides, 2013), in altri in quella cartacea (Bayazit & Askar, 2012).

Le differenze di prestazione emerse in alcuni studi potrebbero dipendere dall'impossibilità di utilizzare le normali strategie (*test-taking strategies*)¹ adottate per i test cartacei (Boevé et al., 2015). Queste strategie comuni di svolgimento del compito, nella maggioranza dei casi, non sono infatti permesse dai test computerizzati, motivo per il quale ancora spesso gli studenti scelgono la modalità cartacea (Hochlehnert Brass, Moeltner & Juenger, 2011) e che potrebbe spiegare le peggiori performance al computer dal momento che è stato verificato che gli esaminandi con più alto successo ai test tradizionali fanno un uso più consistente di queste strategie (Stenlund, Eklöf & Lyrén, 2016) e che il trasferimento delle strategie sviluppate per le verifiche cartacee (ad esempio la gestione del tempo) alle verifiche informatiche non risulta semplice (Walker & Handley, 2016).

Se da alcune ricerche risulta che gli studenti abbiano bisogno del tempo necessario a familiarizzare con i nuovi strumenti elettronici, prima di veder migliorare le loro performance (Deutsch, Herrmann, Frese & Sandholzer, 2012), da altre analisi (Jeong, 2014; Leeson, 2006) non emerge una relazione tra il livello di familiarità con il computer

¹ Le strategie comprendono il segnare le parole chiave, anticipare la risposta alla domanda prima di leggere le alternative, eliminare quelle non corrette, saltare e poi tornare alle domande più difficili marcando le risposte incerte con un punto interrogativo (Stenlund, Eklöf & Lyrén, 2016).

e le performance, e anche quando gli studenti hanno una maggiore esposizione e familiarità con gli strumenti tecnologici, non ottengono migliori risultati ai test. Secondo la review di Leeson (2006) le *caratteristiche degli utenti* – etnia, genere, grado di elaborazione cognitiva, abilità nel dominio specifico, stile di apprendimento, familiarità con il computer, ansia e stress al computer, atteggiamento nei confronti della tecnologia – producono piccole o non significative ripercussioni sulle differenze di prestazioni tra computer-web e paper testing, mentre le potenziali differenze generate *dalle caratteristiche dell'interfaccia* – leggibilità, dimensione e risoluzione dello schermo, scorrimento, dimensione del carattere, lunghezza e spaziatura della linea, quantità di spazio bianco, modalità di presentazione degli item e possibilità di rivedere e modificare le risposte – sembrano essere state ridotte dai recenti avanzamenti tecnologici (Leeson, 2006).

Nel caso di prove computerizzate che prevedano la lettura di lunghi passaggi di testo, la navigazione ipertestuale o test *scenario-based* arricchiti da elementi multimediali ed interattivi, non andrebbero tuttavia sottovalutate alcune criticità emerse dagli studi che hanno analizzato gli effetti della lettura online e su schermo: in questi casi i partecipanti mostrano infatti ancora maggiori difficoltà perché il device digitale può inibire alcune abituali strategie di comprensione (come ad esempio il sottolineare), per la mancanza di ancoraggi spaziali e di importanti elementi informativi di contesto presenti invece nei testi cartacei; per il disorientamento e le maggiori esigenze cognitive richieste dalla gestione di ipertesti, per il sovraccarico cognitivo e l'effetto distraente generato da un eccessivo uso della multimedialità; per lo scorrimento frequente e l'instabilità della pagina digitale che possono ostacolare l'elaborazione di una visione d'insieme dei contenuti; per la scarsa autoregolamentazione dovuta all'eccessiva fiducia nelle proprie prestazioni sullo schermo (Nardi, 2015).

La variabile che sembra incidere in modo più significativo sulle performance è la possibilità offerta dalla valutazione elettronica di ottenere un feedback immediato sul proprio apprendimento. Secondo Hattie e Timperley (2007), che tuttavia non riportano dati riferiti all'apprendimento in contesti online, “è necessario considerare la natura del feedback, i tempi e il modo in cui uno studente ‘riceve’ questo feedback” (p. 101). Anche se il tema è tutt'ora oggetto di dibattito e non vi sono evidenze tali da poter confermare l'una o l'altra ipotesi, sembra che un feedback online immediato (*immediate feedback*) possa produrre migliori prestazioni rispetto allo stesso feedback fornito successivamente (*delayed feedback*) (Leibold & Schwarz, 2015). Le domande a scelta multipla pongono generalmente due ordini di problemi: gli studenti devono scegliere tra un set fisso di risposte senza poter mostrare il loro processo di apprendimento e possono sapere se la risposta fornita è corretta solo in un secondo momento (Merrel, Cirillo, Schwartz & Webb, 2015). Da questo punto di vista un feedback immediato al termine della prova offre l'opportunità di prendere misure immediate per colmare il divario tra l'attuale livello di conoscenza e quello richiesto per superare il test, consente di ridurre lo stress e l'ansia per le verifiche, permette agli studenti di identificare i propri punti di forza e di debolezza autocorreggendosi e fa in modo che gli errori non persistano (JISC, 2010).

3.2. Atteggiamenti e percezioni degli studenti

È stato riscontrato un gap in letteratura tra le numerose ricerche svolte per comparare le differenze di performance tra computer-based e paper-based e quelle che hanno indagato atteggiamenti e preferenze degli studenti (Walker & Handley, 2016). Gli studi condotti in ambito universitario riportano un alto livello di gradimento per le valutazioni basate su computer e web-based, in particolare per la possibilità di ottenere un feedback immediato

(Dermo, 2009; Hillier, 2014; Nardi & Ranieri, in revisione; Ranieri & Nardi, in revisione; Sorensen, 2013). L'utilizzo di computer e mobile-based assessments sembra produrre maggiore motivazione intrinseca, coinvolgimento, migliori atteggiamenti di apprendimento e minor ansia per le prove (per una review si veda Nikou & Economides, 2016). Gli studenti percepiscono l'e-assessments come un elemento di facilitazione dell'apprendimento (Sorensen, 2013), ma per l'accettazione degli esami al computer sembra importante che gli studenti praticino e acquisiscano familiarità con questa nuova modalità di somministrazione del test (Boevé et al., 2015). Una prima esperienza positiva con gli esami elettronici sembra incidere in maniera significativa sulla preferenza espressa per la nuova modalità d'esame (Fluck, Pullen & Harper, 2009; Rout & Patnaik, 2011). La scelta dell'una o dell'altra modalità di test potrebbe essere influenzata dalla tipologia di domande presentate alle verifiche: per le domande aperte gli studenti prediligono gli strumenti tradizionali d'esame (Betlej, 2013), mentre per le domande basate su scenari vi è una larga preferenza espressa per la modalità computer-based (Lim, Ong, Wilder-Smith & Seet, 2006).

Alcuni studenti riportano i vantaggi della scrittura al computer nel migliorare il loro rendimento alle prove (Walker & Handley, 2016), dato che sembra vero in particolare quando si trovano sotto pressione (Fluck et al., 2009). Gli studenti più deboli tendono ad avere un approccio più prudente nei confronti dell'e-assessment rispetto agli studenti più preparati (Fluck et al., 2009; Sorensen, 2013). Non sembra esserci un'associazione positiva tra livello di alfabetizzazione digitale e preferenza espressa per le prove al computer, che sembra invece dipendere in larga parte dalla conoscenza dello strumento tecnologico utilizzato e dalla capacità del singolo studente di adattare le strategie di svolgimento del compito a quelle richieste dall'ambiente di valutazione on-line (Walker & Handley, 2016). In vista dell'accettazione del sistema di valutazione elettronica, l'usabilità e il supporto fornito dal sistema risultano fattori più importanti per le studentesse (Terzis & Economides, 2011), le quali preferiscono utilizzare i dispositivi personali piuttosto che strumenti forniti dall'istituzione (Hillier, 2014). Queste differenze dovute al genere si ridurrebbero tuttavia notevolmente dopo la prima esperienza con i test computerizzati (Deutsch et al., 2012).

3.3. Elementi di maggiore criticità riscontrati dagli studenti

Sintetizzando quanto emerso in letteratura, gli elementi di maggiore criticità della valutazione computerizzata evidenziati dagli studenti riguardano: l'*integrità e la fiducia percepita* (la capacità del sistema elettronico di precludere fenomeni di cheating); l'*affidabilità* (la stabilità delle apparecchiature, della rete e del software), i *problemi tecnici* (ad esempio il timore che il computer possa subire fenomeni di crashing durante l'esame); i *fattori psicologici* (prodotti dallo stress e l'ansia), l'*autoefficacia* (come percezione della propria performance/sicurezza di svolgere il test al computer), la *familiarità* (con l'ambiente e lo strumento di somministrazione del test) (Dermo, 2009; Deutsch et al., 2012; Hillier, 2014); l'*efficienza* (nonostante l'abitudine ad utilizzare i computer per i compiti durante il corso molti studenti scelgono di svolgere l'esame scrivendo a mano per paura di problemi tecnici, per la scarsa capacità di battitura o per il confort prodotto dallo status quo) (Hillier, 2015); i *livelli di attenzione* (rumore generato dalle tastiere e confusione prodotta dalle interfacce utente multi-finestra) (Fluck et al., 2009); l'*equità* (nell'uso di domande pescate in maniera casuale dal database) (Dermo, 2009); il *digital divide* (alcuni studenti sono preoccupati che i colleghi dotati di dispositivi più recenti e con migliori caratteristiche tecniche possano avere un ingiusto vantaggio) (Seow & Soong, 2014).

3.4. Design e comparabilità dei test

Per quanto riguarda la comparabilità dei test svolti al computer, online e su carta le differenze tra i punteggi ottenuti ai test potrebbero essere il risultato proprio della modalità scelta, anche se l'informatizzazione delle domande è stata condotta in modo tale da rendere le versioni delle prove il più possibile comparabili (Buerger, Kroehne & Goldhammer, 2016). L'utilizzo di modalità combinate di somministrazione può dar luogo ai cosiddetti *effetti modalità (mode effects)*: le risposte date in una modalità possono essere significativamente differenti dalle risposte fornite nell'altra, senza la possibilità di stabilire con certezza se tale cambiamento sia *reale* oppure, più semplicemente, costituisca l'esito dell'adozione della nuova modalità. Questi elementi distortivi possono essere ridotti, o talvolta annullati, attraverso l'adozione di un disegno unico (*unified-mode* o *uni-mode design*) che consente di ottenere questionari validi e affidabili (Sala, 2013). Quando si svolgono sperimentazioni in classe, e non in setting di laboratorio, si va necessariamente incontro a fenomeni di distorsione da autoselezione (*self-selection bias*) dal momento che in queste condizioni non è possibile *imporre* una specifica modalità di testing. Esistono tuttavia alcune tecniche statistiche utili a correggere e limitare i possibili bias (Hox, De Leeuw & Zijlmans, 2015). In merito agli aspetti di design i fattori di indagine sono molteplici: i test vanno pensati per la loro fruizione sullo schermo, curando aspetti come il carattere e la grandezza del testo, l'impaginazione, lo scrolling, le funzionalità di feedback, quelle di preview, revisione, consegna e via dicendo (Buerger et al., 2016). Per le implicazioni più strettamente pedagogiche va sicuramente tenuto in considerazione che, in molti casi, la valutazione massiva degli studenti deve necessariamente ricorrere a domande ed esercizi chiusi, tipologia di test preferita dalla valutazione computerizzata, che può portare però con il tempo a focalizzare l'attenzione sulla verifica di un livello superficiale di comprensione e conoscenza (Chalmers & McAusland, 2002). Molti dei test standardizzati sono stati criticati proprio perché prevedono soltanto una valutazione quantitativa dei risultati trascurando la dimensione più qualitativa e per la scarsa capacità che hanno di valutare processi cognitivi di ordine superiore (Fluck et al., 2009), anche se recentemente gli strumenti elettronici si sono evoluti per permettere attività sempre più sofisticate di valutazione (Boyle & Hutchinson, 2009).

3.5. BYOD e possibili sviluppi futuri

Molti paesi stanno mostrando un maggiore interesse per il BYOD ed è possibile trovare sperimentazioni che prevedono l'uso di questo approccio per la valutazione in Australia, Austria, Canada, Danimarca, Finlandia, Germania, Islanda, Norvegia e Singapore (Hillier, 2015). Da alcuni studi sembra che l'utilizzo dei dispositivi personali promuova un apprendimento più profondo, personalizzato e centrato sullo studente, facendo leva sull'attaccamento e la familiarità con il proprio dispositivo (Seow & Soong, 2014). La scelta di utilizzare un device fornito dall'Istituzione, invece degli strumenti personali, potrebbe inoltre avere un impatto negativo sull'accettazione degli esami elettronici da parte degli studenti (Hillier, 2015). Se da un lato il BYOD migliora la scalabilità, permettendo di abbattere molti dei costi necessari all'implementazione di esami elettronici, allo stesso tempo pone nuove sfide alla sicurezza (Dawson, 2015). Le nuove *minacce* derivano dalla possibilità che gli studenti hanno di manomettere il proprio dispositivo (con l'installazione di materiali o strumenti non consentiti) aggirando le funzioni di sicurezza del sistema. Non è detto però che i rischi siano inferiori a quelli di un esame tradizionale e gli strumenti elettronici consentono inedite contromisure agli imbrogli (Sindre & Vegendla, 2015).

Nei prossimi anni assisteremo con molta probabilità al superamento del tradizionale paradigma basato sul testing e all'inizio di una nuova era di *embedded assessment* (Redecker & Johannessen, 2013), dove grazie all'utilizzo di sistemi intelligenti di *assessment analytics* (Ellis, 2013) e alla raccolta, misurazione e analisi dei dati relativi agli allievi e alle loro attività di apprendimento online, sarà possibile interpretare i livelli di competenza, valutare i progressi accademici, prevedere le prestazioni future e adattare i processi valutativi alle esigenze dello studente. La possibilità di integrare elementi multimediali e interattivi, domande adattive, simulazioni, scenari online, ambienti immersivi, realtà virtuale e aumentata, tutor intelligenti, learning analytics, tracciamento comportamentale e data mining, potrà migliorare notevolmente le esperienze valutative e l'obiettività della valutazione rispetto ai reali livelli di preparazione e conoscenza degli esaminandi, ampliando le limitate possibilità offerte dagli esami tradizionali, valutando abilità cognitive e apprendimenti di alto livello e restituendo una valutazione che tenga maggiormente conto della complessità dell'apprendimento e dello stesso processo valutativo (Guàrdia, Crisp & Alsina, 2017).

4. Conclusioni

In questo lavoro abbiamo analizzato la letteratura esistente al fine di indagare il tema della valutazione elettronica nel contesto universitario. Dalla letteratura reperita non sembrano emergere significative problematiche relative alle performance, al gradimento e alle preferenze espresse dagli studenti, rispetto ai tradizionali metodi di valutazione. Emergono inoltre diversi aspetti positivi sotto il profilo organizzativo e gestionale, in relazione all'efficacia e alla semplificazione delle procedure, per quanto riguarda l'obiettività e la qualità del processo valutativo. Il passaggio a modalità e strumenti elettronici di valutazione, come ogni cambiamento, pone inevitabilmente anche nuove sfide e necessita di ripensare gli attuali processi e modelli organizzativi e valutativi. Sembra tuttavia auspicabile che nei prossimi anni, soprattutto in contesti che mostrano ancora un ritardo rispetto al panorama internazionale (come quello Italiano), si riesca a dare un impulso concreto in questa direzione. Secondo l'OECD "gli studenti hanno bisogno di un feedback considerevole, regolare e significativo; gli insegnanti ne hanno bisogno per capire chi sta imparando e come orchestrare il processo di apprendimento" (Dumont, Istance & Benavides, 2010, p. 17) e la valutazione elettronica è in grado di fornire feedback sistematici, tempestivi, ricchi e personalizzati. Sono urgenti ricerche che valutino strumenti e modalità per convertire la vasta mole di dati che in futuro verrà resa disponibile da strumenti sempre più evoluti di *assessment analytics* in un formato significativo che permetta a insegnanti e docenti universitari di utilizzare queste informazioni per ottimizzare il proprio lavoro (Guàrdia et al., 2017). L'analisi dei dati raccolti dall'interazione, dall'assegnazione di compiti e dalla verifica degli studenti in ambienti di apprendimento e piattaforme online di ateneo permetterà di monitorare il livello di apprendimento di un largo numero di studenti, fornire compiti e prove in linea con l'andamento della classe, scoprire specifiche incomprensioni, così come pattern e modalità di svolgimento dei compiti. L'analisi dei processi di apprendimento potrà così servire a migliorare la valutazione e allo stesso tempo l'analisi dei processi valutativi permetterà di comprendere meglio l'apprendimento, in un processo dove i confini tra i due processi diverrà probabilmente sempre più sfumato (Jordan, 2013).

Bibliografia

- Bayazit, A., & Aşkar, P. (2012). Performance and duration differences between online and paper pencil tests. *Asia Pacific Education Review*, 13(2), 219–226.
- Betlej, P. (2013). E-examinations from student's perspective – The future of knowledge evaluation. *Cognition and Creativity Support Systems*, 9, 12–21.
- Boevé, A.J., Meijer, R.R., Albers, C.J., Beetsma, Y., & Bosker, R.J. (2015). Introducing computer-based testing in high-stakes exams in higher education: results of a field experiment. *PLoS ONE*, 10(12), 1–13.
- Boyle, A., & Hutchison, D. (2009). Sophisticated tasks in e-assessment: What are they and what are their benefits? *Assessment & Evaluation in Higher Education*, 34(3), 305–319.
- Buerger, S., Kroehne, U., & Goldhammer, F. (2016). The transition to computer-based testing in large-scale assessments: Investigating (partial) measurement invariance between modes. *Psychological Test and Assessment Modeling*, 58(4), 597–616.
- Calvani, A. (2013). Evidence based (informed?) education: neopositivismo ingenuo o opportunità epistemologica? *Form@re, Open Journal per la Formazione in Rete*, 13(2), 91–101, <http://bit.ly/2CZ76B3> (ver. 15.03.2018).
- Chalmers, D. & McCausland, W.D. (2002). Computer-assisted assessment. In J. Houston & D. Whigham (eds.), *The Handbook for Economics Lecturers: Assessment* (pp. 2–20). Bristol: Economics LTSN.
- Dawson, P. (2015). Five ways to hack and cheat with bring-your-own-device electronic examinations. *British Journal of Educational Technology*, 47(4), 592–600.
- Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203–214.
- Deutsch, T., Herrmann, K., Frese, T., & Sandholzer, H. (2012). Implementing computer-based assessment – A web-based mock examination changes attitudes. *Computers & Education*, 58(4), 1068–1075.
- Dumont, H., Istance, D., & Benavides, F. (Eds.). (2010). *The nature of learning: Using research to inspire practice*. Paris: Educational Research and Innovation, OECD Publishing.
- Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology*, 44(4), 662–664.
- Fluck A, & Hillier, M. (2016). *Innovative assessment with eExams*, Australian Council for Computers in Education Conference, Brisbane. <http://bit.ly/2g7Q5qU> (ver. 18.03.2018).
- Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. *Australasian Journal of Educational Technology*, 25(4), 509–523.
- Guàrdia, L., Crisp, G., & Alsina, I. (2017). Trends and challenges of e-assessment to enhance student learning in Higher Education. In Cano, E. & Ion, G. (eds.), *Innovative Practices for Higher Education Assessment and Measurement* (pp. 36–56). Hershey, PA: IGI Global.

- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review Of Educational Research*, 77(1), 81–112.
- Hillier, M. (2014). The very idea of e-Exams: Student (pre) conceptions. In B. Hegarty, J. McDonald & S.K. Loke (eds.), *Rhetoric and Reality: Critical perspectives on educational technology*. Proceedings ascilite Dunedin 2014 (pp. 77-88). <http://ascilite.org/conferences/dunedin2014/files/fullpapers/91-Hillier.pdf> (ver. 18.03.2018).
- Hillier, M. (2015). e-Exams with student owned devices: student voices. In *Proceedings of the International Mobile Learning Festival 2015: Mobile Learning, MOOCs and 21st Century learning* (pp. 582-608). Hong Kong. http://transformingexams.com/files/Hillier_IMLF2015_full_paper_formatting_fix_ed.pdf (ver. 18.03.2018).
- Hochlehnert, A., Brass, K., Moeltner, A., & Juenger, J. (2011). Does medical students' preference of test format (Computer-based vs. Paper-based) have an Influence on Performance? *BMC Medical Education*, 11(1), 89–95.
- Hox, J.J., de Leeuw, E.D., & Zijlmans, E.A.O. (2015). Measurement equivalence in mixed mode surveys. *Frontiers in Psychology*, 6, 87.
- Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour & Information Technology*, 33(4), 410–422.
- JISC. Joint Information Systems Committee (2010). *Effective Assessment in a Digital Age. A Guide to Technology-Enhanced Assessment and Feedback*. Bristol: Higher Education Funding Council for England. https://facultyinnovate.utexas.edu/sites/default/files/digiassass_eada.pdf (ver. 18.03.2018).
- Jordan, S. (2013). E-assessment: Past, present and future. *New Directions*, 9(1), 87–106.
- Kuikka, M., Kitola, M., & Laakso, M. (2014). Challenges when introducing electronic exam. *Research in Learning Technology*, 22, 22817.
- Leeson, H. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6(1), 1–24.
- Leibold, N., & Schwarz, L.M. (2015). The art of giving online feedback. *Journal of Effective Teaching*, 15(1), 34–46.
- Luecht, R. L., & Sireci (2011). *A review of models for computer-based testing. Research report 2011-2012*. New York, NY: The College Board. Retrieved from <https://files.eric.ed.gov/fulltext/ED562580.pdf> (ver. 18.03.2018).
- Lim, E. C., Ong, B. K., Wilder-Smith, E. P., & Seet R. C. (2006) Computer-based versus pen-and paper testing: students' perception. *Annals of the Academy of Medicine*, 35, 599–603.
- Merrel, J.D., Cirillo, P.F., Schwartz, P.M. & Webb, J.A. (2015). Multiple choice testing using immediate feedback – assessment technique (IF AT®) forms: assessing learning from mistakes. *Higher Education Studies*, 5(5), 50–55.
- Nardi, A. (2015). Lettura digitale vs lettura tradizionale: implicazioni cognitive e stato della ricerca. *Form@re - Open Journal per la Formazione in Rete*, 15(1), 7–29. <http://bit.ly/2xd6Hnb> (ver. 18.03.2018).

- Nardi, A., & Ranieri, M. (in revisione). Comparing Paper-based and BYOD e-text examinations: Impact on students' performance, self-efficacy and satisfaction. *British Journal of Educational Technology*.
- Nikou, S.A., & Economides, A.A. (2013). Student achievement in paper, computer/web and mobile based assessment. *Proceedings of the 6th Balkan Conference on Informatics (BCI)*, Greece. <http://ceur-ws.org/Vol-1036/p107-Nikou.pdf> (ver. 18.03.2018).
- Nikou, S. & Economides, A. (2016). The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement. *Computers in Human Behavior*, 55, 1241–1248.
- Noyes, J.M., & Garland, K.J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352–1375.
- Ranieri, M., & Nardi, A. (in revisione). Su carta o sullo schermo? Studio sulle percezioni delle verifiche digitali in ambito universitario. *TD Tecnologie Didattiche*.
- Redecker, C., & Johannessen, Ø. (2013). Changing Assessment Towards a New Assessment Paradigm Using ICT. *European Journal of Education*, 48(1), 79–96.
- Rout, G. & Patnaik, S. (2011). A case study on e-examination in Universities of Odisha. *International Journal of Internet Computing (IJIC)*, 1(2), 12–20.
- Sala, E. (2013). Le indagini campionarie del nuovo millennio. Quali novità all'orizzonte?. *Quaderni di sociologia*, 57(62), 77–90.
- Seow, T.K., & Soong, S.K.A. (2014). Students' perceptions of BYOD open-book examinations in a large class: a pilot study. In B. Hegarty, J. McDonald & S.K. Loke (eds.), *Rhetoric and Reality: Critical perspectives on educational technology*. Proceedings ascilite Dunedin 2014 (pp. 604–608). <http://ascilite.org/conferences/dunedin2014/files/concisepapers/206-Seow.pdf> (ver. 15.03.2018).
- Sindre, G. & Vegendla, A. (2015). E-exams versus paper exams: a comparative analysis of cheating-related security threats and countermeasures. *Norwegian Information Security Conference (NISK)*, 8(1), 34–45.
- Sorensen, E. (2013). Implementation and student perceptions of e-assessment in a Chemical Engineering module. *European Journal of Engineering Education*, 38(2), 172–185.
- Stenlund, T., Eklöf, H., & Lyrén, P. (2016). Group differences in test-taking behaviour: an example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice*, 24(1), 4–20.
- Terzis, V., & Economides, A. A. (2011). Computer based assessment: Gender differences in perceptions and acceptance. *Computers in Human Behavior*, 27(6), 2108–2122.
- Walker, R., & Handley, Z. (2016). Designing for learner engagement with computer-based testing. *Research in learning technology*, 24, 30083.