

Αξιολόγηση της επίδοσης των μαθητών στην έκθεση μέσω του τεχνολογικού αυτοματισμού

*Βασίλειος Μποζιάρης, Χαράλαμπος Κωνσταντίνου,
Ιωάννης Κωνσταντίνου*

1. Εισαγωγή

Η αξιολόγηση της επίδοσης του μαθητή αποτελεί αναπόσπαστο και σημαντικό μέρος της εκπαιδευτικής διαδικασίας του σχολείου. Συνιστά ένα από τα επίμαχα θέματα που απασχολούν τόσο την εκπαιδευτική όσο και την ερευνητική κοινότητα, αλλά και την ίδια την εκπαιδευτική πολιτική κάθε χώρας. Ειδικότερα, η αξιολόγηση της γραπτής έκφρασης, δηλαδή της έκθεσης, θεωρείται ότι συγκεντρώνει το σύνολο των προβλημάτων που αναφέρονται στο μεθοδολογικό πεδίο της αξιολόγησης της επίδοσης του μαθητή. Η έκθεση, ως μάθημα σε όλες τις εκπαιδευτικές βαθμίδες του σχολείου, συμβάλλει σε καθοριστικό βαθμό στην ανάπτυξη τόσο της γλωσσικής ικανότητας όσο και, συνολικά, της γνωστικής, συναισθηματικής, νοητικής και κοινωνικής ικανότητας του μαθητή. Στη δευτεροβάθμια εκπαίδευση αποκτά ιδιαίτερη βαρύτητα, δεδομένου ότι η αξιολόγησή της, που, κατά κανόνα, αποτυπώνεται με βαθμολογικούς χαρακτηρισμούς, συνδέεται άμεσα με την εισαγωγή των μαθητών στην τριτοβάθμια εκπαίδευση και, επομένως, με τη μελλοντική ακαδημαϊκή και κοινωνική τους εξέλιξη.

Τόσο η βαθμολόγηση όσο και η ανατροφοδότηση, που παρέχονται μέσα από την αξιολόγηση της έκθεσης, αποκτούν βαρύνουσα σημασία για τον μαθητή, από άποψη εκπαιδευτική, επαγγελματική και κοινωνική. Συνεπώς, επιβάλλεται, για λόγους μεθοδολογικούς, η αξιολογική διαδικασία να χαρακτηρίζεται από τις βασικές προϋποθέσεις της έγκυρης, αξιόπιστης και αντικειμενικής αξιολόγησης της επίδοσης των μαθητών, ειδικά στην έκθεση που θεωρείται η πιο πολύπλοκη και δύσκολη περίπτωση αξιολόγησης μαθήματος.

Τα τελευταία χρόνια, έχει αναπτυχθεί μία νέα τεχνική αξιολόγησης, η οποία επιχειρεί να δώσει λύσεις σε βασικά μεθοδολογικά προβλήματα της αξιολόγησης της επίδοσης του μαθητή στην έκθεση. Η νέα αυτή τεχνική ονομάζεται Α.Α.Ε. (Αυτοματοποιημένη Αξιολόγηση της Έκθεσης - Automated Essay Assessment) και εμφανίζεται ως μία καινοτόμος και αποδοτική τεχνολογική πρόταση. Αφορά την αξιολόγηση της έκθεσης από ειδικά λογισμικά που είναι εγκατεστημένα σε ηλεκτρονικούς υπολογιστές. Ιδίως κατά τις τελευταίες

τρεις δεκαετίες έχει κάνει αλματώδη βήματα προς τη βελτίωση της εγκυρότητας, της αντικειμενικότητας και της αξιοπιστίας της αξιολόγησης της επίδοσης, ενώ υπάρχουν πολλοί υποστηρικτές, οι οποίοι την προτιμούν έναντι της παραδοσιακής αξιολόγησης της έκθεσης από τον άνθρωπο. Σε χώρες, όπως οι Ηνωμένες Πολιτείες της Αμερικής, η Αυστραλία και αλλού, χρησιμοποιείται τόσο από τους εκπαιδευτικούς στα σχολεία ως επικουρικό εργαλείο αξιολόγησης και ανατροφοδότησης των εκθέσεων των μαθητών όσο και σε αρκετές επίσημες εξετάσεις, στις οποίες συμμετέχουν μαθητές για την εισαγωγή τους στην τριτοβάθμια εκπαίδευση και στις οποίες ο ένας από τους δύο αξιολογητές είναι η μηχανή, δηλαδή ο υπολογιστής.

Σκοπός, λοιπόν, του παρόντος άρθρου, το οποίο συνιστά ένα από τα πρώτα άρθρα στην ελληνική βιβλιογραφία, είναι να παρουσιάσει βιβλιογραφικά την προβληματική και την επιστημονική συζήτηση γύρω από τη νέα αυτή εναλλακτική τεχνική, επονομαζόμενη αυτοματοποιημένη αξιολόγηση της επίδοσης των μαθητών στην έκθεση. Ειδικότερα, με το άρθρο αυτό παρουσιάζονται οι βασικές πτυχές της συγκεκριμένης τεχνικής, η οποία αφορά την αξιολόγηση του γραπτού λόγου, δηλαδή της έκθεσης των μαθητών, όχι από τον άνθρωπο αλλά από τον ηλεκτρονικό υπολογιστή, όπως είναι αποτυπωμένη στη διεθνή βιβλιογραφία. Επίσης, αναδεικνύονται τα πλεονεκτήματά της, τα οποία έχουν βρει απήχηση σε αρκετούς υποστηρικτές, καθώς και τα μειονεκτήματά της, στα οποία αντιδρούν πολλοί στη χρήση της. Επιπλέον, αξιοποιούνται πορίσματα συγκεκριμένων ερευνών και γίνεται κριτική σύγκριση των δύο τεχνικών, δηλαδή της παραδοσιακής και της αυτοματοποιημένης. Η εργασία ολοκληρώνεται με τις τελικές επισημάνσεις και την προοπτική εφαρμογής της νέας τεχνικής.

Η επιλογή των διεθνών επιστημονικών εργασιών, που αξιοποιήθηκαν στο παρόν άρθρο, βασίστηκε στα ακόλουθα κριτήρια: οι εργασίες να περιγράφουν τα βασικά χαρακτηριστικά των σημαντικότερων συστημάτων Α.Α.Ε., να περιέχουν μελέτες και έρευνες που αξιολογούν την αντικειμενικότητα, την εγκυρότητα και την αξιοπιστία της Α.Α.Ε. ως τεχνικής αξιολόγησης της επίδοσης του μαθητή στην έκθεση, να περιέχουν απόψεις των εκπαιδευτικών και των μαθητών σχετικά με την χρήση των συστημάτων Α.Α.Ε., να περιέχουν τα πλεονεκτήματα και τα μειονεκτήματα της χρήσης της Α.Α.Ε., να περιέχουν, κατά το δυνατόν, τεκμηριωμένες επισημάνσεις και προβλέψεις αναφορικά με την προοπτική της χρήσης της Α.Α.Ε. και, τέλος, να αποτελούν σημαντικές πηγές αναφοράς για μεταγενέστερα άρθρα στη διεθνή βιβλιογραφία.

Η νέα αυτή τεχνική είναι σήμερα ένα ανοικτό ερευνητικό ζήτημα που προσελκύει ερευνητές από διάφορα επιστημονικά πεδία, οι οποίοι επιχειρούν να επιλύσουν μεθοδολογικές διαδικασίες αξιολόγησης που απασχολούν πολύ καιρό την εκπαιδευτική, και όχι μόνο, κοινότητα.

2. Η αξιολόγηση ως έννοια και πρακτική στην εκπαιδευτική διαδικασία

Η «αξιολόγηση» ως έννοια είναι σύμφυτη με τον άνθρωπο ως αναπόσπαστο στοιχείο της ανθρώπινης δραστηριότητας, ενώ αποτελεί μία γνωστή από την αρχαιότητα πρακτική. Για πρώτη φορά έχουν αναφερθεί συστηματοποιημένες διαδικασίες αξιολόγησης του ανθρώπινου δυναμικού στην αρχαία Αίγυπτο και την αρχαία Κίνα, ενώ με το πέρασμα των αιώνων καθιερώθηκε ως αναγκαία διαδικασία στα εκπαιδευτικά συστήματα των χωρών (Srgiven, 1991).

Σε ευρύ επίπεδο δεν υπάρχει τομέας της ανθρώπινης δραστηριότητας από τον οποίο να απουσιάζει η έννοια ή η διαδικασία της αξιολόγησης είτε σε επίσημη (τυπική) είτε σε ανεπίσημη (άτυπη) μορφή, ενώ ως διαδικασία θα συνεχίσει και στο μέλλον να συνοδεύει την πορεία και την εξέλιξη της κοινωνίας σε κάθε τομέα των δραστηριοτήτων του ανθρώπου. Αποτελεί βασικό στάδιο κάθε οργανωμένης και συστηματικής διαδικασίας, η οποία έχει αρχικά προγραμματιστεί και, ακολούθως, εφαρμοστεί, για να διαπιστωθεί ο βαθμός επίτευξης των αρχικά επιδιωκόμενων στόχων και να εντοπιστούν οι παράμετροι που παρεμπόδισαν την επίτευξη αυτών, με απώτερο σκοπό την ανατροφοδότηση και βελτίωση της εκπαιδευτικής δραστηριότητας και των ατόμων που συμμετείχαν σε αυτή (Κωνσταντίνου & Κωνσταντίνου, 2017).

Για τους Κωνσταντίνου & Κωνσταντίνου (2017: 17-19), ως «εκπαιδευτική αξιολόγηση» νοείται «η διαδικασία που αποβλέπει να προσδιορίσει όσο πιο συστηματικά, έγκυρα, αξιόπιστα, αντικειμενικά και δίκαια γίνεται, τη λειτουργικότητα, την αποδοτικότητα και το αποτέλεσμα μίας μαθησιακής και, γενικότερα, εκπαιδευτικής δραστηριότητας, με βάση σαφή κριτήρια, κατάλληλες τεχνικές και συγκεκριμένους σκοπούς, με απώτερο στόχο τη συλλογή πληροφοριών για την ανατροφοδότηση και βελτίωσή της».

Αναφορικά με την αξιολόγηση της επίδοσης του μαθητή, επισημαίνεται ότι αυτή αποτελεί την πλέον διαδεδομένη διαδικασία στην εκπαιδευτική πραγματικότητα και αφορά το σύνολο των διαδικασιών, βάσει των οποίων ελέγχεται ο βαθμός επίτευξης των εκπαιδευτικών στόχων από τον μαθητή, στο πλαίσιο μίας συγκεκριμένης εκπαιδευτικής δραστηριότητας ή ενός εκπαιδευτικού προγράμματος. Η «επίδοση» ως έννοια σχετίζεται με την ποσότητα και την ποιότητα των γνώσεων, ικανοτήτων και δεξιοτήτων, που διαθέτει ο μαθητής σε μία συγκεκριμένη θεματική περιοχή και σε δεδομένη χρονική περίοδο, επηρεαζόμενη από τα ατομικά χαρακτηριστικά του μαθητή, τις επικρατούσες οικογενειακές συνθήκες, το σχολικό και το ευρύτερο κοινωνικό περιβάλλον. Ή, συνοπτικότερα, η δραστηριότητα του μαθητή και το αποτέλεσμά της (Κωνσταντίνου & Κωνσταντίνου, 2017).

Σημειώνεται ότι η Α.Α.Ε., την οποία πραγματεύεται η παρούσα μελέτη, ως μία καινοτόμα και εναλλακτική τεχνική, αφορά το στάδιο επιλογής των

τεχνικών αξιολόγησης της επίδοσης του μαθητή, δηλαδή τη συλλογή πληροφοριών και τον έλεγχο επίτευξης των επιδιωκόμενων εκπαιδευτικών στόχων.

Κυρίαρχο στοιχείο κάθε διαδικασίας αξιολόγησης της επίδοσης αποτελεί η «ερώτηση» που εμπεριέχεται σε όλες τις τεχνικές αξιολόγησης. Ο Δημητρώπουλος (1998: 81-159) διαχωρίζει τις ερωτήσεις σε σύγχρονες (ερωτήσεις πολλαπλής επιλογής, προσαρμογής, σύζευξης, συμπλήρωσης κενού, ιεράρχησης αντικειμένων κ.ά.) και σε παραδοσιακές (ερωτήσεις γλωσσικών μαθημάτων, επίλυσης προβλημάτων και ανάπτυξης). Στις τελευταίες, δηλαδή τις ερωτήσεις ανάπτυξης, οι οποίες αποτελούν βασικό αντικείμενο της παρούσας μελέτης, ο μαθητής καλείται να αναπτύξει ένα κοινωνικό θέμα και να δώσει τις απαιτούμενες απαντήσεις. Πρόκειται για τον πιο γνωστό τρόπο εξέτασης παγκοσμίως, μέσω του οποίου γίνεται ανάπτυξη σύνθετων κοινωνικών θεμάτων, τα οποία απαιτούν οργάνωση σκέψης και έκφρασης, οργάνωση μνήμης και ανάκλησης περιστατικών και φαινομένων, σύνδεση ιδεών, διατύπωση ή χρήση διαδικασιών και αρχών, αξιολόγηση και εφαρμογή γνώσεων, ερμηνεία και ανάλυση φαινομένων κ.ο.κ.

Όπως κάθε αξιολογική διαδικασία, έτσι και στην έκθεση, επιβάλλεται να τηρούνται οι ακόλουθες προϋποθέσεις που εξασφαλίζουν τη μεθοδολογική της πληρότητα:

1. Εγκυρότητα, σύμφωνα με την οποία αξιολογείται αυτό ακριβώς που έχει εξαρχής ορισθεί να αξιολογηθεί, δηλ. ο αρχικά προκαθορισμένος στόχος (δεξιότητες, γνώσεις κ.ά.).
2. Αξιοπιστία, σύμφωνα με την οποία, όσες φορές και αν επαναληφθεί η αξιολογική διαδικασία, κάτω από τις ίδιες συνθήκες, οφείλει να δώσει το ίδιο ή περίπου το ίδιο αποτέλεσμα.
3. Αντικειμενικότητα, σύμφωνα με την οποία διασφαλίζεται ότι η αξιολογική διαδικασία δεν επηρεάζεται από παράγοντες άσχετους προς το αντικείμενο της αξιολόγησης (Κωνσταντίνου & Κωνσταντίνου, 2017: 262-263).

Επίσης, ως χαρακτηριστικά της μεθοδολογικά αποδεκτής ή, αλλιώς, καλής αξιολόγησης, ορίζονται και τα ακόλουθα:

1. Διακριτικότητα, σύμφωνα με την οποία, τα αποτελέσματα της αξιολογικής διαδικασίας πρέπει να κατανέμονται σύμφωνα με μια κανονική ορθότητα (Gaussian).
2. Οικονομία, η οποία σχετίζεται με το σύνολο των πόρων που καταναλώνει η αξιολογική διαδικασία, δηλαδή χρήμα, χρόνο, ανθρωποώρες κ.ά., σε όλα της τα στάδια (εξέταση, διόρθωση γραπτών κ.ά.). Για παράδειγμα, εξετάσεις που απαιτούν τη συγγραφή πολυσέλιδων εκθέσεων, για τη διόρθωση των οποίων απαιτούνται πολλές ανθρωποώρες, έχουν

συνήθως υψηλό κόστος και, έτσι, είναι αντιοικονομικές.

3. Πρακτικότητα, η οποία αναφέρεται στο κατά πόσο μία αξιολογική διαδικασία είναι εύκολη στη χρήση της τόσο για τους αξιολογητές όσο και για τους αξιολογούμενους. Συχνά συμβαδίζει με το κριτήριο της οικονομίας διότι ένα από τα κριτήριά της είναι και η εξοικονόμηση χρόνου και προσωπικού (Κασσωτάκης, 1997: 50-52).
4. Χρηστικότητα, όταν τα αποτελέσματα της αξιολόγησης είναι αξιοποιήσιμα και εκμεταλλεύσιμα από όσους ενδιαφέρονται για αυτά (Αθανασίου, 1985: 25).

Σχετικά με την ανατροφοδότηση, η οποία συνιστά τον απώτερο στόχο της αξιολογικής διαδικασίας, πρόκειται για πρακτική λήψης διορθωτικών, συμπληρωματικών και ενισχυτικών παιδαγωγικών μέτρων από τον εκπαιδευτικό και το σχολείο, τα οποία αφορούν τη βελτίωση τόσο της μαθησιακής διαδικασίας όσο και του μαθητή. Τα μέτρα αυτά συχνά επιβάλλεται να είναι εξατομικευμένα και προσαρμοσμένα στα ιδιαίτερα χαρακτηριστικά κάθε μαθητή. Ο ρόλος της ανατροφοδότησης είναι σημαντικός για την εκπαιδευτική διαδικασία, την ανάπτυξη της προσωπικότητας του μαθητή και, ειδικότερα, τη διευθέτηση δυσκολιών, ελλείψεων, ανεπαρκειών αλλά και δυνατοτήτων του είτε αυτές τον συνοδεύουν από το οικογενειακό, κοινωνικοοικονομικό και πολιτισμικό περιβάλλον του είτε είναι εγγενείς ή επίκτητες (Κωνσταντίνου & Κωνσταντίνου, 2017).

Τέλος, η ανατροφοδότηση μπορεί να βοηθήσει και τον ίδιο τον εκπαιδευτικό να εκτιμήσει την αποτελεσματικότητα της δικής του προσπάθειας και μεθοδολογίας σε κάθε τάξη και σε κάθε μαθητή. Είναι τόσο σημαντική ώστε φτάνει στο σημείο να λογίζεται ως «ίσως το καλύτερο μάθημα για τον εκπαιδευτικό», αρκεί να είναι επιδεκτικός στα μηνύματα που λαμβάνει από τους μαθητές του, έχοντας την πεποίθηση ότι η αυτοβελτίωσή του θα οδηγήσει και στη βελτίωση των μαθητών του (Δημητρόπουλος, 1998: 264-265).

3. Έκθεση και αξιολόγηση

Η επίδοση στον προφορικό και τον γραπτό λόγο είναι από τα βασικότερα κριτήρια της ακαδημαϊκής προόδου ενός μαθητή, ενώ επισημαίνεται και η αιτιώδης σχέση ανάμεσα στη γλωσσική αποτυχία και τη σχολική αποτυχία. Γνώσεις και δεξιότητες που θα μπορούσε να αποκτήσει ο μαθητής στη σχολική του πορεία μπορεί να παραμένουν δυσπρόσιτες λόγω ελλειμμάτων στις θεμελιακές δεξιότητες της γραφής και της ανάγνωσης (Salvia, Ysseldyke & Bolt, 2010).

Σε αναγνώριση της βαρύνουσας σημασίας της γλώσσας για την εξέλιξη των μαθητών, το σχολείο από νωρίς ενδιαφέρθηκε για τη συστηματική καλλιέργεια

και ανάπτυξή της μέσα από ειδικά για τον σκοπό αυτό μαθήματα, τα λεγόμενα γλωσσικά μαθήματα. Η διδασκαλία τους και η αξιολόγηση της επίδοσης των μαθητών σε αυτά αποσκοπεί στο να κάνει τους μαθητές καλούς ακροατές, ομιλητές, αναγνώστες και χειριστές του γραπτού λόγου. Αυτό συμβαίνει ανάλογα με την ηλικία τους, λαμβάνοντας υπόψη τις ιδιαιτερότητες κάθε μαθητή, δηλαδή τις ελλείψεις, τις δυνατότητες, τα ενδιαφέροντα κ.λπ., καθώς και την «εκάστοτε επικοινωνιακή περίσταση», δηλαδή το πλήθος των παραμέτρων που αφορούν τόσο το περιβάλλον εντός του οποίου λαμβάνει χώρα μία γλωσσική - επικοινωνιακή πράξη όσο και τα όσα έχουν υπόψη τους τα άτομα που συνομιλούν, δηλαδή πεποιθήσεις, αξίες, σκοπός, πρόθεση και αποτέλεσμα που επιθυμούν να επιφέρουν κ.λπ. (Αθανασίου, 2003).

Για τον λόγο αυτό, αποκτά ιδιαίτερο ενδιαφέρον η αξιολόγηση του γραπτού λόγου των μαθητών με τη μορφή της ελεύθερης γραπτής έκφρασης ή, αλλιώς, της έκθεσης, ιδίως όταν η επίδοσή τους σε αυτόν μετουσιώνεται στον βαθμό, που λαμβάνουν κυρίως στις εξετάσεις για την εισαγωγή στην τριτοβάθμια εκπαίδευση, ο οποίος είναι καθοριστικής σημασίας για τη μετέπειτα μορφωτική και κοινωνική τους εξέλιξη (Μποζιάρης, 2018).

3.1. Η έκθεση και η προβληματική της αξιολόγησής της

Η αξιολόγηση της επίδοσης του μαθητή στην έκθεση συνδέεται με μία σειρά από προβλήματα και περιορισμούς που επιβαρύνουν τόσο τους μαθητές όσο και τους εκπαιδευτικούς, όπως: α) η συγγραφή της έκθεσης ενδέχεται να καταστεί πηγή απογοήτευσης για εκπαιδευτικούς και μαθητές, β) οι εκπαιδευτικοί δαπανούν αρκετή ενέργεια και χρόνο κατά την αξιολόγηση των εκθέσεων των μαθητών τους, γ) η διαδικασία της αξιολόγησης στερείται εγκυρότητας, αξιοπιστίας και αντικειμενικότητας, με αποτέλεσμα να καθίσταται συχνά και άδικη, δ) κατά την ώρα της συγγραφής αρκετοί μαθητές νιώθουν άβολα και εμφανίζουν ψυχοσωματικές περιπλοκές (π.χ. ιδρώμα χεριών κ.ά.), ενώ βιώνουν και διάφορα αρνητικά συναισθήματα (άγχος, ανησυχία, θλίψη, αγωνία κ.ά.), ε) οι μαθητές, προκειμένου να πάρουν υψηλό βαθμό, συχνά αναγκάζονται να υποκρίνονται, να καταφεύγουν στην αντιγραφή, τον βερμπαλισμό και τον ψευδολογοτεχνισμό ή, ακόμη, οι απαντήσεις τους να μην συνάδουν με τις πεποιθήσεις τους, απλά και μόνο διότι γνωρίζουν ότι θα ηχήσουν ωραία στα αυτιά του εκπαιδευτικού τους, στ) η αξιολόγηση μιας έκθεσης μπορεί να οδηγήσει σε σύγχυση και σε μεγάλες διακυμάνσεις στη βαθμολογία που θα λάβει από διαφορετικούς βαθμολογητές, αν δεν υπάρχουν σαφείς οδηγίες (κριτήρια, κλίμακες κ.ά.) για την αξιολογική διαδικασία ή αν δεν εφαρμοστούν ορθά, ζ) διατυπώνεται αμφιβολία για τον βαθμό με τον οποίο κατατάσσονται οι μαθητές στην αξιολογική διαδικασία, ανάλογα με

την αξία τους και τις ανώτερες πνευματικές και νοητικές τους ικανότητες, ιδίως σε εξετάσεις μεγάλης σημασίας, όπως είναι οι πανελλήνιες εξετάσεις. Ερευνητικά δεδομένα έχουν αναδείξει ότι οι βαθμοί των μαθητών συσσωρεύονται μεταξύ 12 και 15 στη βαθμολογική κλίμακα, γιατί οι αξιολογητές συνδέουν τον βαθμό με τη μορφωτική και κοινωνική εξέλιξη των μαθητών, γεγονός που δείχνει ότι οι βαθμοί δεν χαρακτηρίζονται από διακριτικότητα κ.ά. (Miller, Linn & Gronlund 2009: 242 · Αγγελάκος, 1999: 48-49).

Επιπλέον, υπάρχουν διάφοροι παράγοντες που ασκούν επιρροή στη βαθμολογία μιας έκθεσης, όπως: α) το φαινόμενο του φωτοστέφανου, σύμφωνα με το οποίο ο αξιολογητής τείνει να θεωρεί τις απαντήσεις ενός εξεταζόμενου σωστές ή λάθος ανάλογα με τη γενική εντύπωση της ικανότητας που έχει σχηματίσει για αυτόν, β) η πλάνη της εγγύτητας, σύμφωνα με την οποία η εγγύτητα στον χρόνο, αλλά και στον χώρο, ανάμεσα στην εξέταση και στη βαθμολόγηση επιδρά εξομοιωτικά στην κρίση που διατυπώνει ένας αξιολογητής για έναν αξιολογούμενο, γ) λάθη κοινωνικής αντίληψης, στα οποία υποπίπτει ο εκπαιδευτικός όταν παραβλέπει ή δεν δίνει την ανάλογη βαρύτητα σε λάθη που κάνουν οι «καλοί μαθητές» ή οι μαθητές που προέρχονται από οικογένειες με υψηλότερο κοινωνικό, οικονομικό, μορφωτικό επίπεδο, δ) λάθη στην αξιολόγηση λόγω σειράς του γραπτού κατά τη διαδικασία της διόρθωσης (π.χ. λόγω κόπωσης κ.ά.), ε) τα χαρακτηριστικά του θέματος της έκθεσης και η σαφήνεια στη διατύπωσή του και στ) η γενική εικόνα του γραπτού (καλογραμμένο, κακογραμμένο κ.ο.κ.) κ.ά. (Miller, Linn & Gronlund, 2009: 243-248).

Τέλος, αναφέρεται ότι η επίδραση που μπορεί να ασκήσει στον μαθητή ένα ορθά αξιολογημένο γραπτό είναι καθοριστικής σημασίας για τη βελτίωση της γλωσσικής του ικανότητας (Hyland & Anan, 2006). Επιπλέον, υπάρχει μεθοδολογικό πρόβλημα στην αξιολόγηση της επίδοσης του μαθητή στην έκθεση, δεδομένου ότι είναι ιδιαίτερα δύσκολη η μέτρηση και η αποτίμηση, μέσω ενός βαθμού, των ανώτερων πνευματικών ικανοτήτων και δραστηριοτήτων του ατόμου (Κωνσταντίνου & Κωνσταντίνου, 2017).

4. Η αυτοματοποιημένη αξιολόγηση της έκθεσης μέσω του υπολογιστή ως μία εναλλακτική τεχνική

Διανύουμε μία εποχή, όπου οι Τεχνολογίες της Πληροφορικής και των Επικοινωνιών (Τ.Π.Ε.) έχουν γίνει αναπόσπαστο μέρος της καθημερινότητάς μας. Με την πρόοδο της επιστήμης της Πληροφορικής και τα επιτεύγματα στο επιστημονικό πεδίο της Τεχνητής Νοημοσύνης (Τ.Ν.), οι ερευνητές επιχείρησαν να δώσουν στις μηχανές μια ποιότητα στον τρόπο με τον οποίο λειτουργούν, εφάμιλλη με εκείνη του ανθρώπου.

Οι Τ.Π.Ε. έχουν βρει μία σημαντική θέση και στον τομέα της εκπαίδευσης, γενικότερα, αλλά και, πιο συγκεκριμένα, στη μορφή και το περιεχόμενο της αξιολόγησης. Με τη βοήθειά τους πραγματοποιούνται επίσημες εξετάσεις σε διάφορα σχολεία στις Η.Π.Α., το αποτέλεσμα των οποίων καθορίζει τις αποφάσεις για τη μετέπειτα πορεία της ζωής κάθε μαθητή που συμμετέχει σε αυτές. Οι εξετάσεις αυτές δεν αφορούν μόνο ερωτήσεις κλειστού τύπου, οι οποίες, άλλωστε, είναι πιο εύκολο να αξιολογηθούν από έναν υπολογιστή, αλλά πρόσφατα αφορούν και εξετάσεις στον γραπτό λόγο, στην ανάπτυξη, δηλαδή, θεμάτων με τη μορφή εκθέσεων (Bennett, 2002).

Τελευταία παρατηρείται ένα αυξανόμενο ερευνητικό και εμπορικό ενδιαφέρον για την ανάπτυξη της τεχνικής της Α.Α.Ε., ως μίας εναλλακτικής πρότασης αντί της αξιολόγησης από τον άνθρωπο (Attali, Bridgeman & Trapani, 2010). Ως Α.Α.Ε. νοείται «η αξιολόγηση των εκθέσεων που συντάσσουν οι μαθητές με τη βοήθεια ενός κειμενογράφου, η οποία διενεργείται από κατάλληλο λογισμικό, εγκαταστημένο σε έναν ηλεκτρονικό υπολογιστή» (Scharber, Dexter & Riedel, 2008: 5).

Τα συστήματα Α.Α.Ε. είναι λογισμικά εγκατεστημένα σε έναν ηλεκτρονικό υπολογιστή, τα οποία διαθέτουν κατάλληλα σχεδιασμένους αλγόριθμους και επιχειρούν να αποτιμήσουν και προβλέψουν τη βαθμολογία που θα παρείχε ο βαθμολογητής σε μια έκθεση, συνδέοντάς την με τη σχετική ανατροφοδότηση. Προκειμένου να επιτευχθούν αυτοί οι στόχοι, αρχικά τα λογισμικά αυτά «εκπαιδεύονται» πάνω σε ένα ευμέγεθες σύνολο από εκθέσεις (σύνολο εκπαίδευσης - training set), οι οποίες έχουν ήδη αξιολογηθεί από έμπειρους βαθμολογητές. Ακολούθως, τα λογισμικά αυτά ρυθμίζουν τις διάφορες παραμέτρους τους, προκειμένου να κατασκευάσουν ένα εσωτερικό στατιστικό μοντέλο, το οποίο να είναι ικανό να αναπαράγει τις βαθμολογίες και την ανατροφοδότηση που έχουν λάβει οι εκθέσεις από τους βαθμολογητές. Στην ουσία τα λογισμικά αυτά «μαθαίνουν» να μιμούνται τον τρόπο με τον οποίο ο άνθρωπος αξιολογεί τις εκθέσεις. Μόλις το σύστημα κριθεί ότι έχει «εκπαιδευτεί» επαρκώς, τότε είναι έτοιμο να παρέχει μία αποτίμηση και πρόβλεψη για τη βαθμολογία και την ανατροφοδότηση που θα παρείχαν οι βαθμολογητές σε μία νέα, άγνωστη έκθεση (Ghost & Fatima, 2010).

4.1. Ιστορική αναδρομή

Το 1966 ο Page ανέπτυξε το πρώτο σύστημα Α.Α.Ε. (το P.E.G. - Project Essay Grade), προκειμένου να χρησιμοποιηθεί ως επικουρικό εργαλείο για τη βαθμολόγηση των εκθέσεων των μαθητών της δευτεροβάθμιας εκπαίδευσης στις Η.Π.Α. (Warschauer & Ware, 2006). Το P.E.G. επιχείρησε να εστιάσει στα χα-

ρακτηριστικά του γραπτού λόγου, που είναι εφικτό να εντοπιστούν μέσα σε μία έκθεση από τον ηλεκτρονικό υπολογιστή, δηλαδή στον αριθμό λέξεων, σημείων στίξης, προθέσεων, σπάνιων λέξεων, έκτασης έκθεσης κ.ά., δίνοντάς του τη δυνατότητα να κάνει αποτιμήσεις και προβλέψεις για τη βαθμολογία της έκθεσης. Σύμφωνα με τον Page, τα χαρακτηριστικά αυτά προσεγγίζουν εκείνα του γραπτού λόγου που αξιολογούνται και από τον άνθρωπο (Attali, Bridgeman & Trapani, 2010).

Το 1968 ο Page δημοσίευσε τα αποτελέσματα της έρευνάς του, στα οποία η συσχέτιση μεταξύ των βαθμολογιών των ανθρώπων και του P.E.G., στις ίδιες εκθέσεις, ήταν ιδιαίτερα υψηλή. Επιπλέον, ο Page είχε καταφέρει να αποδείξει ότι ακόμη και έμπειροι αξιολογητές εκθέσεων ήταν δύσκολο να διακρίνουν αν ο βαθμός που παρείχε το P.E.G. στις εκθέσεις των μαθητών ήταν αποτέλεσμα ανθρώπινης ή μηχανικής αξιολόγησης (Wang & Brown, 2007).

Ωστόσο, η καινοτόμος ιδέα του Page προσέκρουσε στα τεχνολογικά εμπόδια της εποχής εκείνης. Συγκεκριμένα, οι H/Y βρίσκονταν στα πρώτα στάδια της εξέλιξής τους με περιορισμένες δυνατότητες, η χρήση τους για την αξιολόγηση των εκθέσεων δεν ήταν συμφέρουσα από οικονομικής και πρακτικής άποψης, καταλάμβαναν μεγάλο όγκο κ.ά., με αποτέλεσμα η πρακτική εφαρμογή τους να μην εμφανίζεται ελκυστική. Ωστόσο, το εγχείρημα του Page συνέβαλε στη διαμόρφωση νέων επιστημονικών προσανατολισμών, αν και χρειάστηκε να περάσουν περίπου δύο δεκαετίες μέχρι να αναζωπυρωθεί το ενδιαφέρον της επιστημονικής κοινότητας για την Α.Α.Ε. (Shermis, 2014).

Προς τα τέλη της δεκαετίας του 1990, τα επιστημονικά πεδία της *Επεξεργασίας Φυσικής Γλώσσας* – Ε.Φ.Γ. (Natural - Language Processing – N.L.P.), της *Ανάκτησης Πληροφορίας* – Α.Π. (Information Retrieval – I.R.) και της *Λανθάνουσας Σημασιολογικής Ανάλυσης* – Λ.Σ.Α. (Latent Semantic Analyses – L.S.A.) έδωσαν στους ερευνητές νέες προοπτικές για τη μέτρηση χαρακτηριστικών στις εκθέσεις των μαθητών, όπως είναι τα συντακτικά, λεξικολογικά και σημασιολογικά φαινόμενα, τα οποία σχετίζονται με πιο άμεσο τρόπο με τα ποιοτικά χαρακτηριστικά του γραπτού λόγου (Attali & Burstein, 2006).

Την περίοδο που ακολούθησε, η έρευνα κινήθηκε προς τις εξής, κυρίως, κατευθύνσεις:

1. Στην τεκμηρίωση της υπόθεσης ότι τα συστήματα Α.Α.Ε. παρέχουν βαθμολογίες ισάξιες με εκείνες των βαθμολογητών (Warschauer & Ware, 2006).
2. Στη διερεύνηση της αξιοπιστίας των συστημάτων Α.Α.Ε., με σκοπό να βελτιώσουν την απόδοσή τους (Powers, Burstein, Chodorow, Fowles & Kukich, 2002β).
3. Στη διερεύνηση της αποτελεσματικής χρήσης των συστημάτων Α.Α.Ε.

μέσα στο σχολείο, ώστε να βοηθήσουν τον εκπαιδευτικό να βελτιώσει την ικανότητα των μαθητών του στο γραπτό τους λόγο (Riedel, Dexter, Scharber & Doering, 2005).

4. Στην κατεύθυνση της εξαγωγής και της αξιολόγησης πιο ποιοτικών χαρακτηριστικών του γραπτού λόγου στις εκθέσεις των μαθητών, τα οποία σχετίζονται με τις ανώτερες πνευματικές ικανότητες του ανθρώπου, και στην παροχή ανατροφοδότησης προς τον μαθητή αλλά και βοήθειας προς τον εκπαιδευτικό (Wang & Brown, 2007).

Το 2012, το φιланθρωπικό ίδρυμα Hewlett Foundation στις Η.Π.Α. διεξήγαγε μία συγκριτική μελέτη για την αποτελεσματικότητα των υφιστάμενων εμπορικών συστημάτων Α.Α.Ε., από όπου προέκυψε το συμπέρασμα ότι πέτυχαν να φτάσουν και σε ορισμένες περιπτώσεις να ξεπεράσουν την ποιότητα, τη συνέπεια και την αξιοπιστία των βαθμολογητών. Η συσχέτιση ανάμεσα στις βαθμολογίες μεταξύ των συστημάτων Α.Α.Ε. και ενός βαθμολογητή βρέθηκε να είναι υψηλότερη από την αντίστοιχη συσχέτιση ανάμεσα στις βαθμολογίες δύο οποιωνδήποτε βαθμολογητών, ενώ μία αυξανόμενη βιβλιογραφική τάση άρχισε να υποστηρίζει πλέον ανοιχτά τα πορίσματα αυτά (Shermis, 2014).

Επίσης, μία ερευνητική ομάδα στην Αυστραλία κατάφερε να πετύχει αυτοματοποιημένες αξιολογήσεις εκθέσεων των μαθητών διαφορετικών τάξεων, οι οποίες, συγκρινόμενες με εκείνες οποιουδήποτε βαθμολογητή, δεν παρουσίαζαν σημαντικές στατιστικές διαφορές σε σχέση με τις αντίστοιχες ανάμεσα σε δύο διαφορετικούς βαθμολογητές. Με άλλα λόγια, η μηχανή, πριν από τα μέσα της προηγούμενης δεκαετίας, έχει φτάσει σε τέτοιο σημείο επιτυχίας, ώστε μπορεί να διακρίνει αν μία έκθεση είναι καλύτερη από μία άλλη, σχεδόν με τόση επιτυχία όσο και ο άνθρωπος (ACARA, 2015).

4.2. Τα κυριότερα συστήματα αυτοματοποιημένης αξιολόγησης της έκθεσης

Στην παράγραφο αυτή παρατίθενται συνοπτικά τα κυριότερα συστήματα Α.Α.Ε., όπως αυτά παρουσιάζονται στη βιβλιογραφία (βλ., επίσης, Μποζιάρης, 2018).

Α. Το e-Rater αναπτύχθηκε το 1999 από μία ερευνητική ομάδα της εταιρείας Educational Testing Service (E.T.S.), με επικεφαλής τον Jill Burstein, προκειμένου να αξιολογεί την επάρκεια στον γραπτό λόγο των υποψήφιων φοιτητών για ακαδημαϊκές μεταπτυχιακές σπουδές στον τομέα της διοίκησης των επιχειρήσεων στα πανεπιστήμια των Η.Π.Α., στο πλαίσιο της εξέτασης που είναι γνωστή ως G.M.A.T. (Graduate Management Admissions Test). Θεωρείται από τα πιο αξιόπιστα και έγκυρα συστήματα Α.Α.Ε. και χρησιμοποιείται σε συνδυασμό με έναν βαθμολογητή στην αξιολόγηση του γραπτού μέρους της εξέτα-

σης G.M.A.T. Η ερευνητική ομάδα ανέπτυξε το e-Rater, εκκινώντας από την υπόθεση ότι οι ανώτερες ποιότητες του γραπτού λόγου συσχετίζονται με άλλα χαρακτηριστικά του, τα οποία μπορούν να εντοπιστούν και να μετρηθούν από τον H/Y. Το e-Rater ήταν σε θέση να βαθμολογεί σε μία έκθεση περισσότερα από 100 χαρακτηριστικά του γραπτού λόγου (συντακτικό, οργάνωση ιδεών, συνάφεια της έκθεσης με το θέμα κ.ά.) και, ακολούθως, μέσω της Βαθμιαίας ή Βηματικής Ανάλυσης Παλινδρόμησης, κατέληγε στην επιλογή ενός τελικού μοντέλου βαθμολόγησης, με το οποίο γινόταν η τελική αποτίμηση για τη βαθμολογία μίας έκθεσης (Attali & Burstein, 2006).

B. Το Intelligent Essay Assessor (I.E.A.) της εταιρείας Pearson Knowledge Technologies (P.K.T.), το οποίο αξιοποιεί τη μέθοδο της Λανθάνουσας Σηματολογικής Ανάλυσης (Latent Semantic Analysis – L.S.A.), η οποία θεωρείται ότι επιτρέπει στον υπολογιστή να μιμηθεί τον τρόπο που αντιλαμβάνεται το νόημα ενός κειμένου ο άνθρωπος εγκέφαλος. Μία έκθεση αναπαρίσταται ως ένας δισδιάστατος πίνακας, κάθε γραμμή του οποίου αντιστοιχεί σε μία και μοναδική λέξη και κάθε στήλη σε ένα μοναδικό νόημα. Κάθε κελί του πίνακα περιέχει μία τιμή που ισούται με το πλήθος των χρήσεων που η λέξη, η οποία αντιστοιχεί στη γραμμή, χρησιμοποιήθηκε στην έκθεση με το νόημα της αντίστοιχης στήλης. Έτσι, κάθε λέξη της έκθεσης αναπαρίσταται ως ένα σημείο σε έναν σημασιολογικό χώρο από έννοιες, δηλαδή με ένα διάνυσμα ίσο με τη γραμμή του πίνακα που αντιστοιχεί στη λέξη αυτή. Με τον τρόπο αυτό, κάθε έκθεση, ως σύνολο από λέξεις, μπορεί να αναπαρασταθεί με ένα πολύ-διάνυσμα σε έναν σημασιολογικό χώρο εννοιών. Το συνολικό νόημα μίας έκθεσης ισούται με το άθροισμα των νοημάτων των λέξεων του κειμένου και μπορεί να μεταβληθεί ριζικά ακόμη και από την τροποποίηση του νοήματος μίας μόνο λέξης (Dikli, 2006).

Η πρόβλεψη του I.E.A. για τη βαθμολογία μίας νέας άγνωστης στο σύστημα έκθεσης είναι περίπου ίση με τη βαθμολογία που έχει λάβει η ήδη βαθμολογημένη έκθεση που απέχει τη μικρότερη απόσταση από τη νέα έκθεση, με εκείνη δηλαδή που είναι πιο κοντά της στο σημασιολογικό χώρο των εννοιών. Χαρακτηριστικά πλεονεκτήματα του I.E.A. είναι ότι: i) εντοπίζει τη λογοκλοπή, κάτι που είναι ιδιαίτερα δύσκολο για έναν βαθμολογητή, ιδίως όταν έχει να αξιολογήσει μεγάλο αριθμό εκθέσεων, ii) εντοπίζει την ομοιότητα ανάμεσα σε δύο εκθέσεις ακόμη και αν η μία έχει προκύψει από την άλλη χρησιμοποιώντας συνώνυμες λέξεις ή παραφράζοντας τις προτάσεις ή αλλάζοντας τη σειρά των προτάσεων κ.ά. (Attali, Bridgeman & Trapani, 2010).

Γ. Το IntelliMetric, της εταιρείας Vantage Learning, για το οποίο πιστεύεται ότι λειτουργεί με τρόπο παρόμοιο με εκείνον που λειτουργεί ο άνθρωπος όταν αξιολογεί μία έκθεση με τη χρήση της μεθόδου της «ολιστικής εντύπωσης».

Στις αρχικές του εκδόσεις, το IntelliMetric μπορούσε να αξιολογεί περισσότερα από τριακόσια (300) σημασιολογικά και συντακτικά χαρακτηριστικά του γραπτού λόγου, πριν αποφανθεί για την τελική βαθμολογία μίας έκθεσης, ενώ σε μεταγενέστερες εκδόσεις του μπορούσε να λάβει υπόψη του περισσότερα από πεντακόσια (500). Λίγα από τα χαρακτηριστικά αυτά είναι γνωστά στην ερευνητική κοινότητα, δεδομένου ότι αποτελεί προϊόν εταιρείας και προστατεύεται από δικαιώματα ευρεσιτεχνίας (Wang & Brown, 2007).

Το IntelliMetric μιμείται τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος κατακτά, αποθηκεύει, ανακαλεί και χρησιμοποιεί τις πληροφορίες. Διαθέτει εσωτερικούς μηχανισμούς ανατροφοδότησης και, με την κατάλληλη καθοδήγηση, είναι δυνατό να «μαθαίνει» από τα ίδια του τα λάθη. Μπορεί εύκολα να χρησιμοποιηθεί για τη βαθμολόγηση εκθέσεων που είναι γραμμένες και σε άλλες γλώσσες, εκτός της αγγλικής (Dikli, 2006). Αποτέλεσε αντικείμενο πειραματισμού και έρευνας πολλών ερευνητών και περισσότερες από εκατόν είκοσι εργασίες έχουν δημοσιευθεί μέχρι το 2007, με παραπλήσια αποτελέσματα, στις οποίες τεκμηριώνεται η εγκυρότητα και η αξιοπιστία του (Wang & Brown, 2007).

Δ. Το Bayesian Essay Test Scoring System (B.E.T.S.Y.), το οποίο στη βιβλιογραφία δεν αποτελεί ένα εμπορικό προϊόν, αλλά προσεγγίζεται περισσότερο ως ένα ερευνητικό εργαλείο (Rudner & Liang, 2002). Το B.E.T.S.Y. στηρίζεται στη θεωρία του Μπέυς (Bayes) για τη δημιουργία ενός συστήματος Α.Α.Ε., το οποίο έχει κοινά στοιχεία με ορισμένες άλλες, υποβοηθητικές τεχνικές, όπως είναι: i) η απόρριψη των άρθρων, των αντωνυμιών και των προθέσεων (feature selection), ii) η αναγωγή όλων των λέξεων της έκθεσης στη ρίζα τους (για παράδειγμα, οι λέξεις μόρφωση, μορφωμένος, μορφώνω, μορφωτικός κ.ά. θα ανάγονταν στη ρίζα μορφω-) πριν την τελική αξιολόγηση της έκθεσης (stop words) και iii) η επιλογή των χαρακτηριστικών εκείνων της έκθεσης, τα οποία είτε πλεονάζουν είτε είναι άσχετα με τα νοήματα της έκθεσης, ενώ η απομάκρυνσή τους από αυτή δεν επηρεάζει την πληροφορία που περιέχει (stemming). Απαιτούνται περίπου 1000 εκθέσεις, προκειμένου το B.E.S.T.Y. να «εκπαιδευτεί», ενώ η βαθμολογία που παρέχει είναι κατηγορική και διαβαθμισμένη, δηλαδή αναδεικνύει αν μία έκθεση επιτυγχάνει / αποτυγχάνει ή αν είναι εξαιρετική / πολύ καλή / καλή / μέτρια κ.ο.κ. Τέλος, έχει υποστηριχθεί ο ισχυρισμός ότι επιτυγχάνει να συνδυάσει τα καλύτερα από τα χαρακτηριστικά των υπόλοιπων συστημάτων Α.Α.Ε. και ότι μπορεί να αξιολογήσει μικρές σε μέγεθος εκθέσεις, οι οποίες πραγματεύονται διάφορα θέματα (Dikli, 2006).

Ε. Το CRASE της εταιρίας Pacific Metrics. Χρησιμοποιεί αλγορίθμους Ε.Φ.Γ. (N.L.P.) για την κατασκευή του μοντέλου βαθμολόγησης και το 2012 διακρίθηκε σε διεθνή Διαγωνισμό Συστημάτων Α.Α.Ε. (Automated Scoring

Assessment Prize Competition - A.S.A.P.), όπου ελέγχθηκε η αποδοτικότητα εννέα (9) συστημάτων A.A.E. σε οκτώ (8) διαφορετικά θέματα. Βρέθηκε να βαθμολογεί τις εκθέσεις το ίδιο αποτελεσματικά ή ακόμη καλύτερα από ό,τι οι έμπειροι βαθμολογητές. Ενδεικτική είναι η έρευνα των Lazendic, Justus & Rabinowitz (2018), η οποία παρείχε τεκμηρίωση για την αξιοπιστία και την εγκυρότητά του, κατά την αξιολόγηση των εκθέσεων των μαθητών στο πλαίσιο της εξέτασης N.A.P.L.A.N. (ετήσιο πρόγραμμα αξιολόγησης των μαθητών που φοιτούν σε σχολεία της Αυστραλίας, στην ανάγνωση – κατανόηση κειμένου, στον γραπτό λόγο και στην αριθμητική). Από την έρευνα προέκυψε ότι το CRASE δεν παραπλανάται κατά την αξιολογική διαδικασία, τουλάχιστον όχι περισσότερο από όσο ο άνθρωπος, ακόμη και αν την παραπλάνηση αυτή προσπαθούν να την μεθοδεύσουν έμπειροι και εξειδικευμένοι βαθμολογητές εκθέσεων, ενώ παρουσιάστηκαν αποδείξεις ότι μπορεί να αναγνωρίσει και να αξιολογήσει, βάσει σαφών κριτηρίων, ακόμη και τα χαρακτηριστικά του γραπτού λόγου που σχετίζονται με τις ανώτερες πνευματικές ικανότητες του ανθρώπου (Lazendic, Justus & Rabinowitz, 2018).

4.3. Επιλεγμένες έρευνες γύρω από τη χρήση συστημάτων A.A.E.

Το μεγαλύτερο μέρος της έρευνας γύρω από την A.A.E. έχει διεξαχθεί από ερευνητές που εργάζονται για τις εταιρίες, στις οποίες αυτά ανήκουν, δεδομένου ότι οι εταιρίες τα προστατεύουν από την ελεύθερη διάθεσή τους, ως περιουσιακά τους στοιχεία. Οι έρευνες αυτές επικεντρώθηκαν, κυρίως, στο να τεκμηριώσουν την εγκυρότητα και την αξιοπιστία των συστημάτων A.A.E. Σε περισσότερες από το 95% των ερευνών αυτών συγκρίνονται οι συντελεστές συσχέτισης ανάμεσα στις βαθμολογίες: α) που παράγουν τα συστήματα A.A.E. με εκείνες των βαθμολογητών και β) που παρέχουν δύο διαφορετικοί βαθμολογητές. Τα αποτελέσματα δείχνουν ότι οι βαθμολογίες των συστημάτων A.A.E. στις εκθέσεις των μαθητών σχεδόν συμφωνούν με εκείνες των βαθμολογητών ή τις προσεγγίζουν περίπου με τον ίδιο τρόπο που τις προσεγγίζουν και οι βαθμολογίες άλλων βαθμολογητών (Warschauer & Ware, 2006).

Ακολουθεί αναφορά στα βασικά ευρήματα αυτής της κατηγορίας ερευνών:

A. Οι Powers, Burstein, Chodorow, Fowles & Kukich (2002α) διαπίστωσαν ότι η ιεραρχική σειρά των αυτοματοποιημένων βαθμολογιών, επί του δείγματος των εκθέσεων της έρευνάς τους, ήταν παρόμοια με την αντίστοιχη των έμπειρων βαθμολογητών, γεγονός που αποτελεί μία περαιτέρω απόδειξη της εγκυρότητας των συστημάτων A.A.E.

B. Οι Powers et al. (2002β) επισημαίνουν ότι είναι πολύ δύσκολο να καταφέρει κανείς να παραπλανήσει τα συστήματα A.A.E., γεγονός που απαιτεί

συγγραφείς με υψηλές δεξιότητες στη χρήση του γραπτού λόγου, αλλά και γνώση του τρόπου με τον οποίο λειτουργούν τα συστήματα αυτά. Τα συστήματα Α.Α.Ε., όμως, αξιολογούν μαθητές, οι οποίοι δεν έχουν τόσο υψηλές δεξιότητες και γνώσεις, ενώ είναι πιο δύσκολο για έναν μαθητή να ρισκιν-δυνεύσει να παραπλανήσει ένα σύστημα Α.Α.Ε. στο πλαίσιο των εξετάσεων, από το αποτέλεσμα των οποίων κρίνεται η μετέπειτα ακαδημαϊκή του πορεία. Το ίδιο επισημαίνουν και οι Shermis & Burnstein (2003, όπ. αναφ. στο Warschauer & Ware, 2006).

Γ. Διάφορες έρευνες έχουν επιχειρήσει να μελετήσουν τη συμπεριφορά των συστημάτων Α.Α.Ε. σε άλλες γλώσσες, πλην της αγγλικής. Οι Cohen, Ben-Simon & Hovan (2003) βρήκαν ότι το IntelliMetric κατάφερε να βαθμολογήσει με επιτυχία εκθέσεις συνταγμένες στα Εβραϊκά. Οι Ishioka & Kameda (2004) έδειξαν ότι το JESS (Automated Japanese Essay Scoring System), το οποίο βαθμολογεί εκθέσεις συνταγμένες στα ιαπωνικά, παρέχει βαθμολογίες, οι οποίες είναι πολύ κοντά στις αντίστοιχες βαθμολογίες που παρέχει το e-Rater στις ίδιες εκθέσεις, οι οποίες έχουν πρώτα μεταφραστεί στην αγγλική γλώσσα.

Δ. Στη συγκριτική έρευνα του Shermis (2014) διερευνήθηκε η εγκυρότητα και η αξιοπιστία εννέα (9) διαφορετικών συστημάτων Α.Α.Ε., με σκοπό να δοθεί απάντηση στο ερώτημα αν αυτά μπορούν να βαθμολογήσουν τις εκθέσεις των μαθητών με τρόπο παρόμοιο με αυτό του ανθρώπου. Το συμπέρασμα ήταν ότι τα συστήματα Α.Α.Ε. έχουν εξελιχθεί σε σημείο που είναι ικανά να ξεπερνούν τα επίπεδα αξιοπιστίας και εγκυρότητας των βαθμολογητών. Οι βαθμολογίες που παρέχουν είναι πολύ κοντά με εκείνες των ανθρώπων, ακόμη και σε απόλυτους αριθμούς.

Ε. Οι Nguyen & Dery (2016) κατασκεύασαν ένα σύστημα Α.Α.Ε., το οποίο κατάφερε να παρέχει αυτοματοποιημένες βαθμολογίες που συσχετίστηκαν (συντελεστής Quadratic Weighted Kappa) κατά 94.5% με τις βαθμολογίες των βαθμολογητών. Δοκίμασαν το σύστημά τους στο ίδιο δείγμα εκθέσεων με εκείνο που είχε χρησιμοποιηθεί στον διαγωνισμό A.S.A.P. το 2012, όπου ο υψηλότερος συντελεστής συσχέτισης που είχε παρατηρηθεί τότε ήταν ίσος με 81% (βλ., επίσης, Shermis, 2014).

Μία άλλη κατηγορία ερευνών αφορά τη χρήση των συστημάτων Α.Α.Ε. στο σχολείο και το κατά πόσο συμβάλλουν στη βελτίωση της ικανότητας του γραπτού λόγου των μαθητών που τα χρησιμοποιούν. Ενδεικτικά αναφέρονται οι ακόλουθες έρευνες:

Α. Οι Elliot & Mikulas (2004) παρατήρησαν ότι κατά τη διάρκεια του σχολικού έτους που διεξήγαγαν την έρευνά τους, όλα τα σχολεία, που χρησιμοποίησαν το σύστημα Α.Α.Ε. MyAccess, βελτίωσαν το Δείκτη Ακαδημαϊκής τους Επίδοσης (Academic Performance Index) περισσότερο από τα προηγού-

μενα έτη, ενώ η μέση επίδοση των μαθητών βελτιώθηκε κατά 40% σε σχέση με την αρχή της έρευνας. Επιπλέον, περίπου το 86% από τους μαθητές, που είχαν χρησιμοποιήσει το MyAccess, ανταποκρίθηκαν με επιτυχία στις Απολυτήριες Εξετάσεις των Γυμνασίων της Καλιφόρνιας (California High School Exit Examination), ενώ το αντίστοιχο ποσοστό για τους μαθητές που δεν το είχαν χρησιμοποιήσει ήταν 46%.

Β. Ο Attali (2004), συγκρίνοντας τις βαθμολογίες της αρχικής με την τελική έκθεση κάθε μαθητή, η οποία προέκυπτε μετά από αναθεώρηση της αρχικής σύμφωνα με την ανατροφοδότηση που παρείχε το σύστημα A.A.E. Criterion, παρατήρησε ότι η συνολική επίδοση των μαθητών που αναθεώρησαν τις εκθέσεις τους βελτιώθηκε από 3,7 (βαθμολογία αρχικής έκθεσης) σε 4,2 (βαθμολογία τελικής αναθεωρημένης έκθεσης), με άριστα το 6. Επιπλέον, οι μαθητές κατάφεραν να μειώσουν κατά 25%, τον ρυθμό με τον οποίο επαναλάμβαναν τα τριάντα (30) πιο συχνά λάθη τους, όπως αυτά τους επισημαίνονταν από το Criterion.

Γ. Ο Grimes (2005, όπ. αναφ. στο Warschauer & Ware, 2006) σημειώνει ότι οι περισσότεροι εκπαιδευτικοί επιλέγουν να εντάξουν τα συστήματα A.A.E. στη διδασκαλία τους. Όμως, συνέχισαν να παρέχουν στους μαθητές τους και τη δική τους ανατροφοδότηση. Ακόμη, κατέληξε στο συμπέρασμα ότι τα συστήματα A.A.E. δεν καταφέρνουν να ασκήσουν τον μαθητή στην ανάπτυξη των ανώτερων πνευματικών ικανοτήτων του στον γραπτό λόγο δίχως την καθοδήγηση και τη βοήθεια του εκπαιδευτικού.

Τέλος, κρίνεται σκόπιμο να γίνει αναφορά στον Les Perelman, πρώην Διευθυντή του προπτυχιακού προγράμματος σπουδών του Τεχνολογικού Ινστιτούτου της Μασαχουσέτης (M.I.T.), ο οποίος λογίζεται ως ο ειδικός που έχει κατανοήσει την εσωτερική λειτουργία και τα αδύνατα σημεία των συστημάτων A.A.E. στον ύψιστο βαθμό. Ενδιαφέρον παρουσιάζει το λογισμικό Babel Generator, που ανέπτυξε με τη βοήθεια προπτυχιακών φοιτητών του. Αυτό συντάσσει μία έκθεση που βασίζεται σε τρεις, το πολύ, λέξεις κλειδιά που του δίνονται ως είσοδο (π.χ. δημοκρατία, ισότητα, πολιτεία). Η έκθεση που συντάσσει δεν έχει κανένα απολύτως νόημα, όταν διαβάζεται από τον άνθρωπο, διότι περιέχει άναρθρες φράσεις. Όμως, τα συστήματα A.A.E. παραπλανούνται από τα εξαιρετικής ποιότητας επιφανειακά χαρακτηριστικά του γραπτού λόγου που περιέχουν οι εκθέσεις που παράγει το λογισμικό αυτό (μέγεθος, σπάνιες λέξεις, ορθά δομημένες φράσεις κ.ά.) και τις βαθμολογούν με υψηλές τιμές, αντί να τις μηδενίσουν (βλ., επίσης, Μποζιάρης, 2018).

Σε άλλες περιπτώσεις συνέταξε εκθέσεις, οι οποίες περιείχαν φτωχούς συλλογισμούς ή ακόμη και παραλογισμούς. Χαρακτηριστικό παράδειγμα αποτελεί και η έκθεση που συνέταξε, ως απάντηση σε ένα θέμα, το οποίο ζητούσε από

τους μαθητές να εξηγήσουν τους λόγους για τους οποίους η φοίτηση στα Πανεπιστήμια της Αμερικής είναι ιδιαίτερα ακριβή. Η έκθεση του Perelman, αν και υποστήριζε το παράλογο επιχείρημα ότι τα δίδακτρα είναι ιδιαίτερα υψηλά, διότι οι βοηθοί των καθηγητών λαμβάνουν μισθό έξι φορές μεγαλύτερο των Πρυτάνεων των Πανεπιστημίων και χρησιμοποιούν συστηματικά τα ιδιωτικά αεροπλάνα των Πανεπιστημίων για να κάνουν πολύ συχνά παραθαλάσσιες διακοπές, έλαβε από το e-Rater βαθμολογία ίση με έξι (6), δηλαδή το μέγιστο της εξάβαθμης κλίμακας (βλ., επίσης, Μποζιάρης, 2018).

Επιπλέον, ο Perelman επιχείρησε να αναπαραγάγει τα αποτελέσματα της έρευνας του Shermis (Shermis, 2014), στην οποία έχει ήδη γίνει αναφορά. Σύμφωνα με τον Perelman, τα ακατέργαστα δεδομένα της έρευνας του Shermis παρείχαν ξεκάθαρες αποδείξεις ότι τα συστήματα Α.Α.Ε. πριμοδότησαν τις εκθέσεις με το μεγαλύτερο μέγεθος. Άρα, είναι υπερβολικά εξαρτημένα από το μέγεθος της έκθεσης, το οποίο, όμως, είναι σημαντικό σε εξετάσεις που διενεργούνται εντός περιορισμένου χρονικού διαστήματος (30-60 λεπτών), όπως στην περίπτωση της έρευνας του Shermis. Με άλλα λόγια, μεγάλες εκθέσεις σε τέτοιου είδους εξετάσεις προλαβαίνουν να συντάσσουν οι πιο καλοί μαθητές και γι' αυτό υπάρχει υψηλή συσχέτιση ανάμεσα στις αυτοματοποιημένες βαθμολογίες και στις βαθμολογίες των βαθμολογητών. Επίσης, αντικρούει και ορισμένα άλλα επιχειρήματα σχετικά με την έρευνα του Shermis (Perelman, 2014).

4.4. Απόψεις των μαθητών και των εκπαιδευτικών για τη χρήση της ΑΑΕ

Στην ενότητα αυτή γίνεται αναφορά στις απόψεις των μαθητών και των εκπαιδευτικών γύρω από τη χρήση της Α.Α.Ε., ένα μέρος έρευνας στο οποίο η επιστημονική κοινότητα δεν έδωσε ιδιαίτερη βαρύτητα εξαρχής:

Α. Οι Scharber, Dexter & Riedel (2008: 15-24) αναφέρουν τα εξής: 1. Όλοι οι μαθητές που συμμετείχαν στην έρευνά τους δεν δίστασαν να χρησιμοποιήσουν το σύστημα Α.Α.Ε. και να πειραματιστούν μαζί του, προκειμένου να βελτιώσουν τη βαθμολογία τους. Έδειξαν περιέργεια για το πώς λειτουργεί ένα σύστημα Α.Α.Ε. και σχεδόν όλοι περιέγραψαν την εμπειρία τους ως ενδιαφέρουσα και διασκεδαστική. 2. Ορισμένοι μαθητές, παρότι υιοθετούσαν την ανατροφοδότηση που τους παρείχε το σύστημα Α.Α.Ε., θεώρησαν ότι δεν αξιολόγησε με ακρίβεια τις νέες εκδόσεις των εκθέσεών τους. Έτσι απογοητεύτηκαν και έχασαν την εμπιστοσύνη τους, ενώ σε επόμενες εκθέσεις δεν χρησιμοποίησαν με τον ίδιο ζήλο την αυτοματοποιημένη ανατροφοδότηση. 3. Οι περισσότεροι από τους μαθητές που έλαβαν χαμηλές βαθμολογίες στις πρώτες εκδόσεις των εκθέσεών τους παρακινήθηκαν πιο έντονα να συντάξουν και νέες βελτιωμένες εκδόσεις των εκθέσεών τους. 4. Πολλοί μαθητές θεώρησαν ότι η

αυτοματοποιημένη ανατροφοδότηση δεν υπήρξε τόσο λεπτομερής και διορατική όσο του δασκάλου τους, όμως όλοι αποδέχτηκαν ότι περιείχε πληροφόρηση ικανή να οδηγήσει σε ποιοτική αναθεώρηση του περιεχομένου των εκθέσεών τους. 5. Αρκετοί μαθητές υποστήριξαν την άποψη ότι τα συστήματα Α.Α.Ε. μπορεί να είναι πραγματικά πιο δίκαια από τον άνθρωπο κατά την βαθμολόγηση. 6. Αρκετοί μαθητές απέφυγαν να πουν ότι εμπιστεύονται την Α.Α.Ε. για την ακρίβειά της, κυρίως ως προς το μέρος της βαθμολόγησης των εκθέσεών τους σε τελικές προαγωγικές εξετάσεις. 7. Οι μισοί σχεδόν μαθητές είπαν ότι επιχείρησαν, κατά κάποιον τρόπο, να «ευχαριστήσουν» το σύστημα Α.Α.Ε., αποδεχόμενοι τις διορθώσεις που τους πρότεινε, προκειμένου να πετύχουν μεγαλύτερες βαθμολογίες, ενώ η όλη διαδικασία τους έκανε πιο ανταγωνιστικούς και ένιωθαν σαν να προσπαθούν να κερδίσουν ένα ηλεκτρονικό παιχνίδι.

Β. Οι Lipnevish & Smith (2009a και 2009b, όπ. αναφ. στο Seifried, Lenhard & Spinath, 2016) αναφέρουν ότι οι μαθητές δήλωσαν να αισθάνονται ανακούφιση και αποφόρτιση από το άγχος της εξέτασης, όταν ήξεραν ότι την έκθεσή τους δεν θα τη διαβάσει άνθρωπος, αλλά η μηχανή και, έτσι, έγραφαν πιο ελεύθερα τις ιδέες τους.

Γ. Οι Seifried, Lenhard & Spinath (2016: 470) αναφέρουν και τα ακόλουθα: 1. Ο βαθμός αποδοχής του αποτελέσματος της αυτοματοποιημένης αξιολόγησης βρέθηκε υψηλότερος στην περίπτωση που οι μαθητές γνώριζαν ότι η έκθεσή τους αξιολογείται από τον άνθρωπο, παρά από τη μηχανή, ανεξάρτητα από το ποιος από τους δύο τη διενεργούσε στην πραγματικότητα. 2. Οι μαθητές προτιμούν οι εκθέσεις τους να αξιολογούνται από τον άνθρωπο.

Δ. Τέλος, οι Shermis, Koch, Page, Keith & Harrington (2002: 17) αναφέρουν ότι οι εκπαιδευτικοί δεν αντιμετωπίζουν αρνητικά τη χρήση των συστημάτων Α.Α.Ε. και θεωρούν ότι μπορεί να τους βοηθήσει να παρέχουν στους μαθητές τους άμεσα πληρέστερη ανατροφοδότηση, κάτι που, συνήθως, δεν προλαβαίνουν να κάνουν ικανοποιητικά. Οι εκπαιδευτικοί έχουν αγγίξει και άλλες πτυχές της χρήσης των συστημάτων Α.Α.Ε., οι οποίες δεν έχουν διερευνηθεί επαρκώς από την επιστημονική κοινότητα, όπως: 1. Τα συστήματα Α.Α.Ε. θα πρέπει να έχουν τη δυνατότητα να επισημάνουν στους μαθητές και τα σημεία της έκθεσής τους, στα οποία τα κατάφεραν σε υψηλό βαθμό. 2. Να μπορούν να παρέχουν εξατομικευμένη ανατροφοδότηση, ανάλογα με τους στόχους που έχουν τεθεί για κάθε μαθητή.

4.5. Η χρήση της ΑΑΕ στην ελληνική εκπαιδευτική πραγματικότητα

Στο πλαίσιο της παρούσας μελέτης δεν κατέστη δυνατό να εντοπιστούν βιβλιογραφικές αναφορές γύρω από την εφαρμογή της Α.Α.Ε. στην ελληνική

εκπαιδευτική πραγματικότητα. Η έρευνα οδηγήθηκε στο Ινστιτούτο Επεξεργασίας Λόγου (Ι.Ε.Λ.), το οποίο λειτουργεί υπό την αιγίδα της Γενικής Γραμματείας Έρευνας και Τεχνολογίας του Υπουργείου Παιδείας και Θρησκευμάτων (Υ.ΠΑΙ.Θ.) και συγκεκριμένα στον κ. Παπαγεωργίου, Χ. επιστημονικά υπεύθυνο του ερευνητικού άξονα της Ε.Φ.Γ. του Ι.Ε.Λ. Προκειμένου να δοθεί η ελευθερία στον ερευνητή του Ι.Ε.Λ. να παρέχει επιστημονικές μαρτυρίες και απόψεις, επιλέχθηκε η τεχνική της μη κατευθυνόμενης συνέντευξης, όπως αυτή περιγράφεται από τον Παρασκευόπουλο (1993), από την οποία προέκυψαν, μεταξύ άλλων, τα ακόλουθα:

Α. Εντός της ελληνικής επικράτειας και υπό την αιγίδα του Υ.ΠΑΙ.Θ. δεν χρησιμοποιείται η Α.Α.Ε. με συστηματικό τρόπο σε καμία βαθμίδα εκπαίδευσης, ενώ δεν αξιοποιείται η Α.Α.Ε. ούτε από άλλους φορείς (π.χ. G.R.E., T.O.E.F.L. κ.ά.).

Β. Στο παρελθόν, υπήρξε απροθυμία από το Υ.ΠΑΙ.Θ. να παραχωρήσει τις γραπτές εκθέσεις των μαθητών που συμμετείχαν στις Πανελλαδικές Εξετάσεις, ώστε να αποτελέσουν πηγή για έρευνα ή ακόμη και για την εκπαίδευση ενός συστήματος Α.Α.Ε.

Γ. Το Ι.Ε.Λ. είχε υποβάλει ερευνητική πρόταση στο Υ.ΠΑΙ.Θ. για τη χρηματοδότηση της ανάπτυξης ενός συστήματος Α.Α.Ε. στην ελληνική γλώσσα. Ωστόσο, παραμένει ακόμη σε εκκρεμότητα.

Δ. Δεν είναι γνωστή κάποια ολοκληρωμένη ερευνητική προσέγγιση για την εφαρμογή της Α.Α.Ε. στην ελληνική γλώσσα, όμως υπάρχουν κάποιες περιστασιακές και ασύνδετες μεταξύ τους βιβλιογραφικές προσεγγίσεις, οι οποίες σε συνδυασμό με τα ερευνητικά πορίσματα για την εφαρμογή της Α.Α.Ε. στην αγγλική γλώσσα, θα μπορούσαν να αποτελέσουν τη βάση για τη δημιουργία ενός συστήματος Α.Α.Ε. και στην ελληνική γλώσσα: i) το Ι.Ε.Λ. έχει παρουσιάσει κάποια πορίσματα στο Διεθνές Συνέδριο Γλωσσολογίας στον τομέα της αξιολόγησης της γνώσης της ελληνικής γλώσσας και τον τομέα της κατανόησης κειμένου στην ελληνική γλώσσα, ii) ο Μικρός, Γ. από το Καποδιστριακό Πανεπιστήμιο, έχει εργαστεί πάνω στην υφολογία και στην ποσοτική ανάλυση της ελληνικής γλώσσας, iii) η Τζεβελέκου, Μ. από το Ι.Ε.Λ. έχει ασχοληθεί με παραμέτρους της αξιολόγησης της ελληνικής γλώσσας σε πληθυσμούς μαθητών που προέρχονται από τη Θράκη, iv) με την κατανόηση μίας έκθεσης στην ελληνική γλώσσα έχουν ασχοληθεί και ξένοι οργανισμοί (π.χ. Πανεπιστήμιο της Στουτγκάρδης στη Γερμανία).

4.6. Κριτική και απολογισμός από τη χρήση της ΑΑΕ

Οι υποστηρικτές των συστημάτων Α.Α.Ε. χρησιμοποιούν ως επιχείρημα υπέρ της αξιοποίησής τους τα αποτελέσματα των ποικίλων ερευνών, τα οποία ανα-

γνωρίζουν την υψηλή εγκυρότητα, αντικειμενικότητα, αξιοπιστία, οικονομικότητα και χρηστικότητα τους. Ωστόσο, οι περισσότερες από τις έρευνες έχουν διεξαχθεί από τις εταιρίες που έχουν αναπτύξει τα συστήματα Α.Α.Ε., οι οποίες, ενδεχομένως, προσπαθούν να προασπίσουν τα οικονομικά τους συμφέροντα και, έτσι, ίσως δημοσιεύουν ελλιπή ή ακόμη και μεροληπτικά ερευνητικά αποτελέσματα (Wang & Brown, 2007).

Από την άλλη πλευρά υπάρχουν κι εκείνοι που αντιδρούν και εναντιώνονται στη χρήση τους, εκφράζοντας εύλογες ανησυχίες για την εγκυρότητα και την αξιοπιστία των συστημάτων αυτών, ενώ στον πυρήνα της διαμάχης ανάμεσα στις δύο θέσεις, έχει βρεθεί το θέμα του κατά πόσο τα συστήματα Α.Α.Ε. μπορούν να υποκαταστήσουν τον άνθρωπο στη διαδικασία της αξιολόγησης της πολυδιάστατης ποιότητας του γραπτού λόγου. Επιπλέον, επικαλούνται το γεγονός ότι αυτά επιδοκιμάζουν την ικανότητα των μαθητών να ανταποκρίνονται σε επιφανειακά χαρακτηριστικά του γραπτού λόγου που είναι μετρήσιμα από τον Η/Υ, παρά στο να καταστούν ικανοί να παράγουν σαφή νοήματα και να αναπτύξουν τη δημιουργικότητα, τη φαντασία και την ιδιαίτερη προσωπικότητά τους (Wang & Brown, 2007).

Από τα ευρήματα των σχετικών μελετών προκύπτουν τα ακόλουθα πιο σημαντικά πλεονεκτήματα και μειονεκτήματα των συστημάτων Α.Α.Ε.:

Πλεονεκτήματα

Τα αυτοματοποιημένα συστήματα μπορούν άμεσα να βαθμολογήσουν μία έκθεση, επιτρέποντας στον εκπαιδευτικό να τα χρησιμοποιεί υποβοηθητικά και να αξιολογεί μεγάλο αριθμό εκθέσεων, παρέχοντας την ανάλογη εξατομικευμένη ανατροφοδότηση σε κάθε μαθητή του. Αυτό βρίσκει άμεση εφαρμογή σε εξετάσεις όπου απαιτείται η διόρθωση χιλιάδων εκθέσεων σε πολύ σύντομο χρόνο (π.χ. G.R.E., T.O.E.F.L., G.M.A.T. κ.ά.), πλεονεκτώντας σε χρόνο και χρήμα. Επιπλέον, η Α.Α.Ε. θεωρείται πιο αξιόπιστη και αποτελεσματική, διότι παρέχει την ίδια βαθμολογία για την ίδια έκθεση οποιαδήποτε στιγμή και αν την βαθμολογήσει. Επίσης, βαθμολογεί όλες τις εκθέσεις με τα ίδια κριτήρια, εξασφαλίζοντας ίση μεταχείριση στους μαθητές, η οποία δεν μπορεί να διασφαλιστεί στην περίπτωση εξετάσεων, όπου τα γραπτά βαθμολογούνται από διαφορετικούς βαθμολογητές (Dikli, 2006).

Βαθμολογούν τις εκθέσεις των μαθητών ισάξια με τους διαφορετικούς βαθμολογητές, ενώ, επιπλέον, δεν επηρεάζονται από παράγοντες που μπορεί να κάνουν τον άνθρωπο να βαθμολογεί με μειωμένη αντικειμενικότητα, αξιοπιστία και εγκυρότητα, όπως είναι η κόπωση, η απροσεξία, η απόσπαση της προσοχής κ.ά. (Weigle, 2013).

Αρκετά από τα συστήματα Α.Α.Ε. επιτρέπουν στους μαθητές να αποθηκεύουν προσωρινά όλες τις εκδόσεις των εκθέσεων που συντάσσουν μαζί με τις βαθμολογίες και τις οδηγίες ανατροφοδότησης για κάθε μία από αυτές, κάτι το οποίο συμβάλλει στην εφαρμογή της διαμορφωτικής αξιολόγησης από τον εκπαιδευτικό (Weigle, 2013).

Μειονεκτήματα

Τα συστήματα Α.Α.Ε. δεν είναι ικανά να αξιολογήσουν τα ποιοτικά χαρακτηριστικά μίας έκθεσης με τον τρόπο που γίνεται από τον άνθρωπο (Weigle, 2013). Έχουν κατηγορηθεί, μάλιστα, ότι αφαιρούν το ανθρώπινο στοιχείο κατά την αξιολόγηση του γραπτού λόγου στις εκθέσεις των μαθητών και, συνεπώς, υποβαθμίζουν την ανθρώπινη διάσταση της αλληλεπίδρασης ανάμεσα στον μαθητή και τον εκπαιδευτικό (Dikli, 2006).

Οι μηχανές δεν μπορούν να εντοπίσουν σε μία έκθεση τα ουσιώδη χαρακτηριστικά του ανθρώπινου λόγου, όπως είναι η συλλογιστική, η επάρκεια των αποδεικτικών στοιχείων, η καλή αίσθηση, η ηθική στάση, η πειστική επιχειρηματολογία, η οργάνωση, η σαφήνεια, η ακρίβεια του λόγου κ.ά. ή, ακόμη, τις νύξεις ή τις αναφορές που γίνονται σε γνωστά γεγονότα και σε ρήσεις μεγάλων προσωπικοτήτων, σε γνωστή βιβλιογραφία, στο χιούμορ κ.ά. Αντί αυτών, αξιολογούν μόνο τα επιφανειακά χαρακτηριστικά, όπως το μέγεθος κειμένου, το λεξιλόγιο κ.ά., τα οποία τυγχάνει να συσχετίζονται υψηλά με τη ποιότητα της παραγωγής του γραπτού λόγου που μπορεί να αντιληφθεί μόνο ο άνθρωπος. Επιπλέον, εμφανίζουν μία ευπάθεια σε νέες μορφές εξαπάτησης από πλευράς των αξιολογούμενων που σχετίζονται με τα μέσα που διαθέτει η τεχνολογία. Άλλοτε εντοπίζουν λάθη που δεν υπάρχουν και αγνοούν λάθη που υπάρχουν. Δεν συσχετίζουν μία έκθεση με την ατομική πρόοδο που παρατηρείται σε έναν συγκεκριμένο μαθητή, ενώ αδικούν τις εκθέσεις μαθητών που προέρχονται από μειονοτικές ομάδες ή έχουν την αγγλική γλώσσα ως δεύτερη γλώσσα. Τέλος, οι εταιρείες, τηρώντας μυστικοπάθεια, δεν επιτρέπουν τη διεξαγωγή ανεξάρτητων ερευνών για τα προϊόντα τους, ώστε να δοθούν αντικειμενικές απαντήσεις σε ερωτήματα, όπως τα παραπάνω (Human Readers, 2018· Wang & Brown, 2007).

Σε εξετάσεις μεγάλης κλίμακας, όπου με βάση τη βαθμολογία ενός υποψηφίου σε μία γραπτή έκθεση λαμβάνονται αποφάσεις πολύ σημαντικές για την ακαδημαϊκή και κοινωνική εξέλιξή του, δεν είναι δίκαιο και κοινωνικά αποδεκτό οι αποφάσεις αυτές να λαμβάνονται από μία μηχανή. Αλλά ακόμη και στην τάξη, η αξιολόγηση του γραπτού λόγου είναι ανάγκη να γίνεται από τον εκπαιδευτικό, ο οποίος γνωρίζει τους μαθητές και τις ιδιαιτερότητές τους, έχει

τον έλεγχο του προγράμματος σπουδών και της διδασκόμενης ύλης και μπορεί να κατανοεί τους σκοπούς και τη χρησιμότητα της αξιολογικής διαδικασίας, όποια μορφή και αν λάβει αυτή είτε διαμορφωτική είτε τελική (Weigle, 2013).

Αν ο μαθητής έχει κάνει κάποιο λάθος σε μία φράση είναι δύσκολο για τη μηχανή να συναγάγει το νόημα που επιθυμούσε να αποδώσει ο μαθητής, αξιολογώντας τα νοήματα των φράσεων που προηγούνται ή έπονται της φράσης που περιέχει το λάθος (Weigle, 2013).

Τα λάθη που αλλοιώνουν εντελώς το νόημα μιας φράσης θεωρούνται και τα πιο σοβαρά. Το σύστημα Α.Α.Ε. ενδέχεται να μην εντοπίσει το λάθος, διότι η πρόταση είναι συντακτικά σωστή και, επιπλέον, ενδέχεται να μην έχει συναντήσει ανάλογο παράδειγμα κατά τη φάση της εκπαίδευσής του (Weigle, 2013).

Ορισμένοι ερευνητές επιφυλάσσουνται ακόμη και για τα μηνύματα ανατροφοδότησης των συστημάτων Α.Α.Ε. προς τους μαθητές, γιατί, όπως υποστηρίζουν, αυτά θα κάνουν τους μαθητές να προσαρμόσουν τον γραπτό τους λόγο προς την κατεύθυνση επιδίωξης υψηλότερων βαθμολογιών από τη μηχανή και όχι απαραίτητα προς την κατεύθυνση της βελτίωσης της ποιότητας του γραπτού τους λόγου, όπως την αντιλαμβάνεται ο άνθρωπος. Με την Α.Α.Ε. οι μαθητές απευθυνόμενοι σε μηχανές οδηγούνται σε έναν φορμαλιστικό λόγο με αμφίβολη επίδραση ή αντίκτυπο στους συνανθρώπους τους (Wang & Brown, 2007).

Άλλοι προβληματισμοί

Τον Μάρτη του 2006, τα ΜΜΕ στις Η.Π.Α. δημοσίευσαν ότι, εξαιτίας ενός μηχανικού λάθους, εκθέσεις, που έγραψαν χιλιάδες μαθητές στο πλαίσιο της εξέτασης SAT, βαθμολογήθηκαν με λάθος τρόπο. Δύο εβδομάδες αργότερα ανακοινώθηκε επίσημα ότι 4400 εκθέσεις έλαβαν βαθμολογίες χαμηλότερες από όσο άξιζαν και 613 υψηλότερες. Όμως, πολλοί ήταν εκείνοι, οι οποίοι δεν πίστεψαν αυτές τις επίσημες ανακοινώσεις. Ακολούθησαν ακόμη ορισμένες αστοχίες από τη χρήση της Α.Α.Ε. και, έτσι, στις 12 Μαρτίου του 2013, μια ομάδα εκπαιδευτικών, που αυτοαποκαλούνται «Επαγγελματίες Εναντίον της Μηχανικής Βαθμολόγησης των Μαθητικών Δοκιμιών», κυκλοφόρησε ένα διαδικτυακό υπόμνημα κατά των συστημάτων Α.Α.Ε., το οποίο συγκέντρωσε άμεσα χιλιάδες υπογραφές, μεταξύ των οποίων και ο Noam Chomsky (Human Readers, 2018).

Δεν έχει τεκμηριωθεί ότι τα αυτοματοποιημένα συστήματα ανταποκρίνονται επαρκώς στους εξατομικευμένους στόχους που έχει οριστεί να πετύχει κάθε μαθητής μέσα από την εκπαιδευτική πρακτική στην τάξη, αλλά και στους γενικότερους στόχους που έχουν τεθεί από την εκπαιδευτική πολιτική της χώρας όπου χρησιμοποιούνται τα συστήματα αυτά (Shermis, 2014).

Επομένως, είναι χρήσιμο να διερευνηθούν περαιτέρω οι τομείς του γραπτού λόγου στους οποίους βελτιώθηκαν οι μαθητές που χρησιμοποίησαν Α.Α.Ε. (επιφανειακοί ή πιο ουσιώδεις), καθώς και με ποιον τρόπο επήλθε η βελτίωση αυτή (Warschauer & Ware, 2006).

Περαιτέρω διερεύνησης χρήζει και το κατά πόσο είναι εφικτό να τροποποιηθεί η συμπεριφορά ενός μαθητή που συντάσσει μία έκθεση, όταν γνωρίζει ότι αυτή πρόκειται να αξιολογηθεί από ένα σύστημα Α.Α.Ε. Αν είναι γνωστά εκ των προτέρων τα χαρακτηριστικά του γραπτού λόγου που αξιολογούνται από το σύστημα, τότε οι μαθητές, με την κατάλληλη καθοδήγηση, μπορούν να προσαρμόσουν το ύφος της γραφής τους, ώστε να το παραπλανήσουν και να επιτύχουν μεγαλύτερες βαθμολογίες (Shermis, 2014).

Σημαντικό είναι και το ζήτημα της κατάρτισης των εκπαιδευτικών στη χρήση των συστημάτων Α.Α.Ε., ώστε να μπορούν να εκμεταλλευτούν τα πλεονεκτήματα που αποφέρει η χρήση τους, τόσο προς τους μαθητές τους όσο και προς τους ίδιους (Weigle, 2013).

Πολλά σχολεία, προκειμένου να συμμετάσχουν σε έρευνες σχετικά με τη χρήση συστημάτων Α.Α.Ε., χρειάστηκε να προμηθευτούν περισσότερους Η/Υ, αυξάνοντας με αυτόν τον τρόπο την αναλογία των Η/Υ ανά μαθητή. Όμως, δεν διερευνήθηκε επαρκώς αν η ενίσχυση της επίδοσης των μαθητών οφειλόταν αποκλειστικά στη χρήση της Α.Α.Ε. ή στην αύξηση της αναλογίας των Η/Υ ανά μαθητή (Warschauer & Ware, 2006).

Όπως έχει ήδη αναφερθεί, τα συστήματα Α.Α.Ε. εντοπίζουν και αξιολογούν διαφορετικά χαρακτηριστικά του γραπτού λόγου από ό,τι ο άνθρωπος, ενώ επισημαίνονται υψηλές συσχετίσεις ανάμεσα σε αυτές τις δύο ομάδες χαρακτηριστικών. Ωστόσο, οι συσχετίσεις αυτές δεν έχουν διερευνηθεί επαρκώς και, έτσι, δεν είναι γνωστή και επιστημονικά τεκμηριωμένη η αναλογία και η εξάρτηση ενός χαρακτηριστικού του γραπτού λόγου της μίας ομάδας σε σχέση με τα χαρακτηριστικά της άλλης ομάδας (Weigle, 2013).

5. Τελικές επισημάνσεις και προοπτική της αυτοματοποιημένης αξιολόγησης

Οι τάσεις και οι προοπτικές στις Η.Π.Α. αλλά και σε άλλες χώρες καταδεικνύουν ότι η Α.Α.Ε. εισχωρεί όλο και περισσότερο στις διαδικασίες, μέσα από τις οποίες γίνεται η επαγγελματική και κοινωνική επιλογή των μαθητών. Ήδη διατυπώνονται ισχυρισμοί ότι η Α.Α.Ε. και η παραδοσιακή αξιολόγηση της έκθεσης εμφανίζουν δυνατότητες αλληλοσυμπλήρωσης και ότι συνδυάζοντας τα πλεονεκτήματα και των δύο είναι εφικτό να ενισχυθεί η αποδοτικότητα και η αποτελεσματικότητα της αξιολόγησης, καθώς

και η ποιότητα της διενεργούμενης αξιολογικής διαδικασίας της έκθεσης (Xi, 2010).

Η Weigle (2013: 50) αναφέρει ότι «παρά τις όποιες φωνές υπέρ της κατάρτησης ή της μείωσης της χρήσης των συστημάτων Α.Α.Ε. για την αξιολόγηση και τη διδασκαλία της γλωσσικής ικανότητας των μαθητών, θα ήταν αφελές να υποστηρίξει κανείς απόλυτα ότι θα χρησιμοποιούνται όλο και λιγότερο στο μέλλον για τη βαθμολόγηση των εκθέσεων των μαθητών, αλλά και για την παροχή ανατροφοδότησης προς αυτούς». Το βάρος θα πρέπει να δοθεί στη συνέχιση της προσπάθειας για βελτίωση των συστημάτων αυτών.

Η υιοθέτηση κάθε τεχνολογικής καινοτομίας στην εκπαιδευτική πρακτική οδηγεί στη σταδιακή αναδιαμόρφωση της διδακτικής και της μαθησιακής διαδικασίας και συνοδεύεται από πλεονεκτήματα και μειονεκτήματα. Η Α.Α.Ε. έχει δώσει λύσεις σε ορισμένα από τα προβλήματα που σχετίζονται με την αξιολόγηση της γλωσσικής ικανότητας των μαθητών, κυρίως στο πλαίσιο εξεταστικών δοκιμασιών που έχουν ως βάση την έκθεση. Ωστόσο, όπως και κάθε άλλη εκπαιδευτική καινοτομία, δεν έχει φέρει τις «επιθυμητές» λύσεις στα πολύπλοκα προβλήματα που αντιμετωπίζει η εκπαίδευση και ιδίως η αξιολόγηση της επίδοσης του μαθητή (Warschauer & Ware, 2006). Υποστηρίζεται ότι οι εκπαιδευτικοί επιβάλλεται να δείξουν προθυμία στην υιοθέτηση της νέας τεχνολογικής πρότασης, να εκπαιδευτούν κατάλληλα και να υποστηριχθούν κατά τη χρήση της και να μην την αντιμετωπίζουν ως κάτι που έρχεται για να δυσκολέψει το έργο τους (Weigle, 2013).

Ένα ακόμη θέμα που είναι ανάγκη να διερευνηθεί στο μέλλον σχετίζεται με το μέγεθος των εκθέσεων που βαθμολογούν τα συστήματα Α.Α.Ε. Τα συστήματα Α.Α.Ε. επιτυγχάνουν να βαθμολογούν με μεγάλη εγκυρότητα και αξιοπιστία εκθέσεις που έχουν μικρά μεγέθη (μεταξύ πεντακοσίων -500- και χιλίων -1000- λέξεων). Ωστόσο, δεν έχουν διεξαχθεί έρευνες που να υποστηρίζουν την υπόθεση ότι μπορούν να τα καταφέρουν με την ίδια επιτυχία και με μεγαλύτερες εκθέσεις, αλλά και με άλλες μορφές του γραπτού λόγου, όπως είναι η λογοτεχνία και η ποίηση, μολονότι διατυπώνεται έντονη επιφύλαξη σχετικά με το αν μπορούν να τα καταφέρουν σε τέτοιες μορφές δοκιμίων (Shermis et al., 2002).

Όλα τα μειονεκτήματα της Α.Α.Ε., όπως αυτά αναφέρονται στη βιβλιογραφία, αναμένεται ότι τα επόμενα χρόνια θα προσελκύσουν όλο και περισσότερους ερευνητές, όχι μόνο από την Πληροφορική αλλά και από άλλα επιστημονικά πεδία (γλωσσολόγους, μαθηματικούς, ψυχολογολόγους, φιλοσόφους, ανθρωπολόγους, νευροεπιστήμονες κ.ά.), οι οποίοι θα παρέχουν όλο και πιο ευφύεις λύσεις, βελτιώνοντάς την διαρκώς (Xi, 2010). Με την έννοια αυτή, έχουν προκύψει ήδη νέα επιστημονικά πεδία, όπως: Αυτοματοποιημέ-

νη Αξιολόγηση Σύντομων Γραπτών Απαντήσεων (Automated Short-Answer Scoring), Αυτοματοποιημένη Αξιολόγηση της Ποιότητας και της Συνοχής του Λόγου (Automated Evaluation of Discourse Coherence Quality), Αυτοματοποιημένη Ανάλυση των Συναισθημάτων κατά την Αξιολόγηση των Εκθέσεων (Automated Sentiment Analysis for Essay Evaluation), Ανίχνευση και Ερμηνεία της Κίνησης των Οφθαλμών του Αναγνώστη κ.ά. (Weigle, 2013).

Συνεπώς, τα συστήματα Α.Α.Ε. θα πρέπει βελτιωθούν και να εξελιχθούν τα επόμενα χρόνια και προς άλλες κατευθύνσεις, από τις οποίες είναι ασύνδετα και αποκομμένα στην παρούσα φάση εξέλιξής τους, όπως: α) Να αποθηκεύουν και να λαμβάνουν υπόψη τους πληροφορίες που παρέχονται και από άλλες τεχνικές αξιολόγησης (συστηματική παρατήρηση, πορτφόλιο, ρουμπρίκα, παιδαγωγικό ημερολόγιο κ.ά.), καθώς και πληροφορίες που αφορούν τις οδηγίες προς τον μαθητή, οι οποίες έχουν προκύψει ως ανατροφοδότηση είτε από τα ίδια τα συστήματα Α.Α.Ε. είτε από τον εκπαιδευτικό, αλλά και το κατά πόσο εφαρμόζονται από τον μαθητή. β) Να επιτρέπουν στον εκπαιδευτικό να τα προγραμματίζει κατάλληλα, ορίζοντας σαφείς εξατομικευμένους διδακτικούς και παιδαγωγικούς στόχους για κάθε μαθητή. γ) Να λαμβάνουν υπόψη ακόμη και τους στόχους που έχει θέσει η σχολική μονάδα, ανάλογα με τα ιδιαίτερα χαρακτηριστικά του σχολείου και της γεωγραφικής περιοχής του. δ) Τέλος, να παρέχουν τη βαθμολογία και την ανατροφοδότηση για κάθε έκθεση του μαθητή, ανάλογα με τις πληροφορίες αυτές.

Όμως, ακόμη και στην περίπτωση που θα κατασκευαστούν στο μέλλον λογισμικά ικανά να εκπαιδεύσουν και να αξιολογήσουν τον μαθητή στη γλωσσική του ικανότητα καλύτερα από ό,τι ο εκπαιδευτικός, ο μείζον και δικαιολογημένος προβληματισμός συνεχίζει να υφίσταται: Η μηχανή θα μπορέσει να συμβάλει ώστε ο μαθητής να βασιστεί στις δικές του ατομικές δυνατότητες και ιδιαιτερότητες και στις ανθρώπινες και κοινωνικές αξίες; Δηλαδή, θα του επιτρέψει, ειδικότερα, να αναπτύξει την αυτενέργεια στη σκέψη και τις ενέργειές του, στη μορφωτική και κοινωνική του συμπεριφορά και στη διασφάλιση των προϋποθέσεων ανάπτυξης των δυνατοτήτων του; Οι άνθρωποι στο μέλλον θα συγγράφουν πνευματικά έργα που θα διαφέρουν σημαντικά από εκείνα των υπόλοιπων συνανθρώπων τους και θα φέρουν την ταυτότητα του συγγραφέα τους ή οδηγούμαστε προς μία εποχή, όπου όλοι οι άνθρωποι θα εκπαιδεύονται από τη μηχανή και έτσι οι απόψεις τους θα μοιάζουν πολύ μεταξύ τους; Μήπως το όραμα αυτό δεν είναι ακόμη ώριμο ή μήπως είναι πολύ μακριά η εποχή αυτή, στην οποία θα είναι λίγοι εκείνοι που θα ξεχωρίζουν μέσα από τον λόγο τους, τις σκέψεις τους, τη δράση τους, τις πρωτοβουλίες τους κ.ο.κ., μέσα από τον ιδιαίτερο τρόπο που υποστηρίζουν και αρθρώνουν την ατομικότητά τους;

Αναφορικά με τους προβληματισμούς αυτούς, μπορεί κανείς να κάνει διάφορες εκτιμήσεις για το μέλλον, βασιζόμενος στις απόψεις σπουδαίων ερευνητών, στοχαστών και τεχνοκρατών της εποχής μας. Πράγματι, ο τομέας της Τ.Ν. παρουσιάζει συνεχώς νέα επιτεύγματα, δημιουργώντας μηχανές που προσομοιάζουν όλο και πιο πιστά με τον τρόπο που λειτουργεί ο άνθρωπος. Ο βρετανός Καθηγητής Stephen Hawking υποστήριξε τη θέση ότι οι προσπάθειες του ανθρώπου να δημιουργήσουν «σκεπτόμενες μηχανές» θέτουν κινδύνους στην ίδια του την ύπαρξη: «Η ανάπτυξη της πλήρους τεχνητής νοημοσύνης θα μπορούσε να σημάνει το τέλος της ανθρώπινης φυλής ... έπειτα από κάποιο σημείο εξέλιξης της Τ.Ν., οι μηχανές θα μπορούσαν να αναλάβουν τον έλεγχο της εξέλιξής τους και να ανασχεδιάσουν τη δομή, τη λειτουργία και την τελειοποίησή τους, χωρίς να κάνουν λάθη, αλλά και με πολύ πιο γρήγορους ρυθμούς, από εκείνους της βιολογικής εξέλιξης του ανθρώπου, οπότε ο άνθρωπος δεν θα μπορέσει να τις ανταγωνιστεί» (BBC News, 2014).

Συμπληρώνοντας τις σκέψεις αυτές, θα υποστηρίζαμε ως κατακλείδα ότι το μέλλον είναι αυτό που, μέσω των σχετικών επιστημονικών ερευνών και της περαιτέρω εξέλιξης της τεχνολογίας, θα μπορέσει να αναδείξει και να τεκμηριώσει τις προσδοκώμενες απαντήσεις στους συγκεκριμένους προβληματισμούς. Αναμφίβολα, τα επιτεύγματα της τεχνολογίας είναι πολύ σημαντικά και επωφελή για τον άνθρωπο και την κοινωνία, στον βαθμό, όμως, που δεν θα αγνοούνται ή θα περιθωριοποιούνται τα βασικά και ιδιαίτερα χαρακτηριστικά τους, τα οποία σχετίζονται με την ίδια την υπόσταση, τις οικουμενικές αξίες και την προοπτική τους.

Βιβλιογραφία

- ACARA (Australian Curriculum Assessment and Reporting Authority) NASOP (National Assessment and Surveys Online Program) Research Team (2015). *An evaluation of automated scoring of NAPLAN persuasive writing*. ACARA. Retrieved 23 December, 2020, from http://nap.edu.au/_resources/20151130_ACARA_research_paper_on_online_automated_scoring.pdf.
- Αγγελάκος, Κ. (1999). *Η αξιολόγηση της έκθεσης στις γενικές εξετάσεις* (Αδημοσίευτη διδακτορική διατριβή). Πανεπιστήμιο Ιωαννίνων, Φιλοσοφική Σχολή, Τμήμα Φιλοσοφίας-Παιδαγωγικής-Ψυχολογίας, Ιωάννινα.
- Αθανασίου, Λ. (1985). *Η αξιολόγηση της «έκθεσης» στο λύκειο (ψυχοπαιδαγωγική προσέγγιση)*. Αθήνα: Εκπαιδευτήρια Δούκα.
- Αθανασίου, Λ. (2003). *Γλώσσα - γλωσσική επικοινωνία και διδασκαλία στην*

- πρωτοβάθμια και δευτεροβάθμια εκπαίδευση. Ιωάννινα.
- Attali, Y. (2004, April). *Exploring the feedback and revision features of Criterion*. Paper presented at the National Council on Measurement in Education conference, San Diego, CA.
- Attali, Y., Bridgeman, B., & Trapani, C. (2010). Performance of a Generic Approach in Automated Essay Scoring. *Journal of Technology, Learning, and Assessment*, 10(3).
- Attali, Y., & Burstein, J. (2006). Automated Essay Scoring With e-rater V.2. *Journal of Technology, Learning, and Assessment*, 4(3).
- BBC News. (2014). Stephen Hawking warns artificial intelligence could end mankind.
- Bennett, E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment (JTLA)*, 1(1).
- Cohen, Y., Ben-Simon, A., & Hovav, M. (2003, October). *The effect of specific language features on the complexity of systems for automated essay scoring*. Paper presented at the International Association of Educational Assessment Annual Conference, Manchester.
- Δημητρώπουλος, Ε. (1998). *Εκπαιδευτική αξιολόγηση. Η αξιολόγηση του μαθητή. Θεωρία - πράξη - προβλήματα*. Αθήνα: Γρηγόρης.
- Dikli, S. (2006). An Overview of Automated Scoring of Essays. *Journal of Technology, Learning, and Assessment*, 5(1).
- Elliot, S., & Mikulas, C. (2004, April). *The impact of MY Access! use on student writing performance: a technology overview and four studies*. Paper presented at the annual meeting of the American Educational Research Association, San Diego.
- Ghost, S., & Fatima, S. S. (2010). Design of an automated essay grading (AEG) system in Indian Context. *International Journal of Computer Applications*, 1(11), 71-76. doi: 10.5120/237-391
- Human Readers. (2018). *Professionals Against Machine Scoring Of Student Essays In High-Stakes Assessment*. Retrieved 20 December, 2020, from <http://humanreaders.org/petition/>
- Ishioka, T., & Kameda, M. (2004, January). *Automated Japanese essay scoring system: jess*. Paper presented at the Proceedings of the 15th International Workshop on Database and Expert Systems Applications. Zaragosa, Spain, 4-8.
- Κασσωτάκης, Μ. (1997). *Η αξιολόγηση της επιδόσεως των μαθητών. Μέσα, μέθοδοι, προβλήματα, προοπτικές*. Αθήνα: Γρηγόρης.
- Κωνσταντίνου, Χ., & Κωνσταντίνου, Ι. (2017). *Η αξιολόγηση στην εκπαί-*

- δευση. *Η αξιολόγηση του εκπαιδευτικού έργου, του εκπαιδευτικού και του μαθητή ως θεωρία και πράξη*. Αθήνα: Gutenberg.
- Lazendic, G., Justus, J. A., & Rabinowitz, S. (2018). *NAPLAN online automated scoring research program: research report*. ACARA (Australian Curriculum Assessment and Reporting Authority) Research Team.
- Miller, M. D., Linn, L. R., & Gronlund, E. N. (2009). *Measurement and Assessment in Teaching* (10th edition). New Jersey: Pearson Education Inc.
- Μποζιάρης, Β. (2018). *Η αυτοματοποιημένη αξιολόγηση της επίδοσης του μαθητή στην έκθεση ως μια νέα εναλλακτική τεχνική* (Αδημοσίευτη μεταπτυχιακή διατριβή). Πανεπιστήμιο Ιωαννίνων, Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Ιωάννινα.
- Nguyen, H., & Dery, L. M. (2016, November). *Neural Networks for Automated Essay Grading*. The 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016).
- Παρασκευόπουλος, Ι. (1993). *Μεθοδολογία επιστημονικής έρευνας*. Τόμος Β'. Αθήνα.
- Perelman, L. (2014). When “the state of the art” is counting words. *Assessing Writing*, 21, 104-111.
- Powers, D. E., Burstein, J. C., Chodorow, M. S., Fowles, M. E., & Kukich, K. (2002α). Comparing the validity of automated and human scoring of essays. *Journal of Educational Computing Research*, 26(4), 407-425.
- Powers, D. E., Burstein, J. C., Chodorow, M. S., Fowles, M. E., & Kukich, K. (2002β). Stumping e-rater: challenging the validity of automated essay scoring. *Computers in Human Behavior*, 18, 103-34.
- Rudner, L. M., & Liang, T. (2002). Automated Essay Scoring Using Bayes' Theorem. *Journal of Technology, Learning, and Assessment*, 1(2).
- Riedel, E., Dexter, S. L., Scharber, C., & Doering, A. (2005). Experimental evidence on the effectiveness of automated essay scoring in teacher education cases. *Journal of Educational Computing Research*, 35(3).
- Salvia, S. J., Ysseldyke, E. J., & Bolt, S. (2010). *Assessment in special and inclusive education* (11th edition). Boston, MA: Wadsworth/Cengage Publications.
- Scharber, C., Dexter, S., & Riedel, E. (2008). Students' Experiences with an Automated Essay Scorer. *Journal of Technology, Learning, and Assessment*, 7(1).
- Srciven, M. S. (1991). *Evaluation thesaurus* (4th ed.). California: Sage publications.
- Seifried, E., Lenhard, W., & Spinath, B. (2016). Automated essay assessment: Effects on students' acceptance and on learning – related characteristics.

- Psihologija*, 49(4), 469-482.
- Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing*, 20, 53–76.
- Shermis, M. D., Koch, C. M., Page, E. B., Keith, T. Z., & Harrington, S. (2002). Trait ratings for automated essay grading. *Educational and Psychological Measurement*, 62(1), 5-18.
- Wang, J., & Brown, M., S. (2007). Automated Essay Scoring Versus Human Scoring: A Comparative Study. *The Journal of Technology, Learning, and Assessment*, 6(2).
- Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10(2), 1-24.
- Weigle, C. S. (2013). English as a second language writing and automated essay evaluation. In M.D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 36-54). New York: Routledge.
- Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, 27(3), 291-300.

Abstract

The present paper attempts to bibliographically cover the Automated Essay Assessment, a particular emerging assessment technique, which has recently introduced as a technological alternative to the traditional student essays' assessment technique, conducted by humans. Many software programs or platforms have recently been developed, which are underpinning the educational process and used for the assessment of essays in large scale examinations. The paper starts by presenting the topic of the evaluation in education, followed by analyzing the significant importance of essay writing in the process of developing linguistic, cognitive, mental and social characteristics to the students. The topic of the traditional assessment of students' performance in the free written expression, as it is performed by human, is presented as well, as well as the topic of Automated Essay Assessment, as it is described in the international literature. Automated Essay Assessment platforms present many advantages and have won many supporters, but, on the other hand, have also cause several objections and many adversaries have raised serious arguments against their usage. The paper, finally, concludes by listing the most important findings that occur from the critical comparison between the traditional and the automated essay assessment.

Key-words: Assessment, student performance, essay assessment, automated essay assessment

Βασίλειος Μποζιάρης

MSc Παν. Ιωαννίνων, ΠΕ86

Τηλ.: 6939250702

E-mail: vboziari@yahoo.gr

Χαράλαμπος Κωνσταντίνου

Ομ. Καθηγητής Π.Τ.Δ.Ε. Παν. Ιωαννίνων

Τηλ.: 6977392181

E-mail: chkonsta@uoi.gr

Ιωάννης Κωνσταντίνου

Δρ. Παν. Ιωαννίνων, ΠΕ70

Τηλ.: 6944291635

E-mail: jeankonstantinou@gmail.com