

Η αξιοπιστία και η εγκυρότητα στην αξιολόγηση της γραπτής έκφρασης από εκπαιδευτικούς

ΙΩΑΝΝΗΣ Κ. ΔΗΜΑΚΟΣ¹

ΠΕΡΙΛΗΨΗ

Σκοπός της εργασίας είναι να παρουσιάσει μια εναλλακτική πρόταση στα σταθμισμένα, διαγνωστικά-ανιχνευτικά εργαλεία μαθησιακών ικανοτήτων της γραπτής έκφρασης. Διεθνώς η πρόταση *teachers as tests*, δηλαδή οι εκπαιδευτικοί ως κριτήρια αξιολόγησης, υποστηρίζεται από πολλές έρευνες σε μια σειρά γνωστικών αντικειμένων και σε πολλά εκπαιδευτικά συστήματα διάφορων χωρών. Στην εργασία γίνεται μια σύντομη ανασκόπηση αυτών των διεθνών ερευνών και στη συνέχεια παρουσιάζονται δεδομένα από το ελληνικό εκπαιδευτικό σύστημα. Τρία διαφορετικά σετ δεδομένων αναλύονται. Συνολικά, μαθητές των τριών τελευταίων τάξεων του δημοτικού έγραψαν εκθέσεις διάφορων θεμάτων, οι οποίες αξιολογήθηκαν (όχι από τους εκπαιδευτικούς των μαθητών) με μια σειρά κριτηρίων, όπως σύνολο γραπτών λέξεων, σύνολο ορθογραφημένων λέξεων, ορθές αλληλουχίες λέξεων, μονάδες ιδεών, αριθμός ρημάτων, παραγράφων κ.λπ. Στη συνέχεια οι εκπαιδευτικοί των μαθητών αξιολόγησαν τις ικανότητες των μαθητών στην ανάγνωση, την ορθογραφία, τη γραπτή έκφραση σε κλίμακα Likert 5 σημείων (1=πολύ χαμηλή, 5=πολύ υψηλή). Τα δεδομένα έδειξαν ότι οι κρίσεις των εκπαιδευτικών είχαν υψηλή εσωτερική συνέπεια (βάσει του δείκτη *Cronbach's a*) και υψηλή, στατιστικώς σημαντική συνάφεια με τα κριτήρια αξιολόγησης των γραπτών κειμένων των μαθητών. Αν και περισσότερες έρευνες χρειάζονται για να επαληθευτούν περαιτέρω τα δεδομένα της παρούσας έρευνας, ωστόσο, η συστηματική και ενεργός συμμετοχή των εκπαιδευτικών στη μαθησιακή αξιολόγηση των ικανοτήτων και προβλημάτων των μαθητών είναι εφικτή και ίσως να μπορεί να προσφέρει δεδομένα και πληροφορίες που δεν συλλέγονται κατά τη διάρκεια της ψυχομετρικής αξιολόγησης των μαθητών.

Λέξεις-κλειδιά: Εκπαιδευτικοί, Αξιολόγηση, Κριτήρια

1. Εισαγωγή

Ο ρόλος του εκπαιδευτικού στη διαδικασία διαγνωστικής αξιολόγησης και παρέμβασης στην περίπτωση των μαθητών με ειδικές εκπαιδευτικές ανάγκες έχει μια αρκετά μεγάλη ιστορία που

φθάνει μέχρι τις πρώτες ημέρες της Ψυχομετρίας. Για τη μελέτη της εγκυρότητας της κλίμακας νοημοσύνης του Binet χρησιμοποιήθηκε δείγμα εκπαιδευτικών της εποχής. Όταν ο Terman στάθμισε την κλίμακα νοημοσύνης Stanford-Binet, βασίστηκε σε εκπαιδευτικούς της περιο-

1. Διεύθυνση: Ιωάννης Δημάκος, Πανεπιστήμιο Πατρών, Σχολή Κοινωνικών και Ανθρωπιστικών Επιστημών, Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, 26500, Πατρα. e-mail: idimakos@upatras.gr.

χής του για να του υποδείξουν τόσο μαθητές με μέση ευφυΐα όσο και μαθητές και στα δύο άκρα της κατανομής (Terman & Childs, 1912). Ο Binet και ο Terman αναγνώρισαν την ικανότητα των εκπαιδευτικών να διακρίνουν με ακρίβεια μαθητές με μέσο επίπεδο ικανότητας και να τους διαφοροποιήσουν με επιτυχία από τους συμμαθητές τους, των οποίων η επίδοση ήταν άνω ή κάτω του μετρίου.

Σε αντίθεση, λοιπόν, με τις πρώιμες ψυχομετρικές πρακτικές όπου η κρίση του εκπαιδευτικού ήταν το εξωτερικό κριτήριο εγκυρότητας για τα υπό κατασκευή εργαλεία αξιολόγησης, η τακτική που ακολουθήθηκε στις επόμενες δεκαετίες ήταν διαφορετική. Οι αρχικές κρίσεις των εκπαιδευτικών για κάποιον μαθητή με πιθανές μαθησιακές δυσκολίες ή άλλες εκπαιδευτικές ανάγκες εθεωρούντο μη έγκυρες. Αν η αρχική κρίση του εκπαιδευτικού για τον εν λόγω μαθητή ήταν τελικώς ορθή (με άλλα λόγια, αν το αποτέλεσμα της διαδικασίας αξιολόγησης ήταν η διάγνωση μαθησιακών δυσκολιών ή κάποιας άλλης ειδικής εκπαιδευτικής ανάγκης), τότε ήταν το ψυχομετρικό τεστ που εθεωρείτο έγκυρο και αξιόπιστο και όχι ο εκπαιδευτικός που είχε αρχικά αναγνωρίσει το πρόβλημα του μαθητή (Gresham, Reschly & Carey, 1987).

Συχνά οι κρίσεις των εκπαιδευτικών προέρχονται από κάποια κλίμακα ή/και ερωτηματολόγιο. Λόγου χάρη, γονείς και εκπαιδευτικοί συχνά συμπληρώνουν κλίμακες συμπεριφοράς που αφορούν στην ενδοσχολική συμπεριφορά ή στην αντίστοιχη συμπεριφορά του μαθητή στο σπίτι, όπως οι κλίμακες Achenbach που απευθύνονται σε γονείς, εκπαιδευτικούς ή στην περίπτωση εφήβων μαθητών στους ίδιους τους μαθητές (Achenbach, 1991. Achenbach & Rescorla, 2002). Τέτοιες κλίμακες δίνουν στον ειδικό που ασχολείται με τον εν λόγω μαθητή μια λεπτομερή εικόνα της γενικότερης συμπεριφοράς του μαθητή σε διάφορα περιβάλλοντα. Εκτός από τις κλίμακες που σχετίζονται με τη συμπεριφορά των μαθητών, ο Hoge μελέτησε τις ψυχομετρικές ιδιότητες των κλιμάκων που χρησιμοποιούν οι εκπαιδευτικοί για την αξιολόγηση των ακαδημαϊκών

δεξιοτήτων των μαθητών αυτών (1983). Σύμφωνα με τη μελέτη, οι τότε διαθέσιμες κλίμακες ήταν δύο ειδών: είτε εξειδικευμένες σε κάποιο γνωστικό αντικείμενο είτε αφορούσαν τη γενικότερη σχολική επίδοση του μαθητή. Εντούτοις, όλες οι κλίμακες διέθεταν επαρκή εγκυρότητα κριτηρίου (συγκρινόμενες με σταθμισμένα τεστ ακαδημαϊκών επιδόσεων), ικανοποιητική εγκυρότητα περιεχομένου και αξιοπιστία επαναλαμβανόμενων μετρήσεων. Σε μεταγενέστερη επισκόπηση της σχετικής βιβλιογραφίας από τον ίδιο ερευνητή, επισημάνθηκε και πάλι η ανάγκη οι αξιολογικές κρίσεις, που συχνά χρησιμοποιούνται ως ψυχομετρικά εργαλεία να διέπονται από τις ίδιες ψυχομετρικές ιδιότητες που ισχύουν για τα εργαλεία αυτά (Hoge & Coladarsi, 1989). Στην ίδια επισκόπηση ερευνών, οι δείκτες συνάφειας μεταξύ αξιολογικών κρίσεων των εκπαιδευτικών και της επίδοσης των μαθητών σε κάποια τεστ σχολικής επίδοσης κυμάνθηκαν από 0,28 έως 0,92 (διάμεση τιμή 0,66), γεγονός που, σύμφωνα με τους συγγραφείς, «... υποδηλώνει μέτρια προς ισχυρή αντιστοιχία μεταξύ κρίσεων εκπαιδευτικών και μαθητικής επίδοσης» (Hoge & Coladarsi, 1989, σελ. 303) και ενισχύει την εγκυρότητα της χρήσης των αξιολογικών κρίσεων των εκπαιδευτικών.

Οι DuPaul, Rapport και Perriello μελέτησαν τη δυνατότητα εφαρμογής τέτοιων κλιμάκων στη διαδικασία διαγνωστικής αξιολόγησης μαθητών με προβλήματα συμπεριφοράς (1991). Οι εν λόγω ερευνητές παρουσίασαν την κλίμακα APRS (Academic Performance Rating Scale), ένα εργαλείο που χρησιμοποιούσε τις κρίσεις των εκπαιδευτικών για να αξιολογήσει την ακαδημαϊκή επίδοση μαθητών. Η κλίμακα APRS περιείχε 19 ερωτήματα που κάλυπταν διάφορες πτυχές της σχολικής εργασίας των μαθητών (μαθηματικά, γλώσσα, ανάγνωση, γραπτή έκφραση) αλλά και της ικανότητας των μαθητών να ακολουθούν τις οδηγίες των εκπαιδευτικών, να αφομοιώνουν νέες πληροφορίες και να ολοκληρώνουν δραστηριότητες που έχει αναθέσει ο εκπαιδευτικός της τάξης. Οι ερευνητές ανέφεραν τα εξής ψυχομετρικά χαρακτηριστικά για την κλίμακα αυτής: πολύ

υψηλή αξιοπιστία επαναλαμβανόμενων μετρήσεων ($r=0,95$) και εσωτερική συνέπεια (συντελεστής *Cronbach's a*=0,95). Επίσης, η κλίμακα είχε και την απαιτούμενη διακριτική ικανότητα, καθώς μπορούσε να διαφοροποιήσει δυο ομάδες μαθητών, τη μία που αποτελείτο από μαθητές που παρουσίαζαν συμπτώματα ΔΕΠΥ (Διαταραχή Ελλειμματικής Προσοχής Υπερκινητικότητας) και την άλλη που περιλάμβανε μαθητές με «κανονική» συμπεριφορά.

Μια άλλη κρίσιμη εφαρμογή τέτοιων κλιμάκων που βασίζονται στις αξιολογικές κρίσεις των εκπαιδευτικών είναι και η δυνατότητα αναγνώρισης μαθητών με μαθησιακές δυσκολίες (ΜΔ). Οι Gresham και συνεργάτες παρουσίασαν την κλίμακα TRAP (Teacher Rating of Academic Performance), μία κλίμακα αξιολόγησης για την αναγνώριση μαθητών με ΜΔ (Gresham, Reschly & Carey, 1987). Η κλίμακα TRAP περιείχε μόνο πέντε (5) ερωτήματα τύπου Likert τα οποία συνέκριναν την επίδοση των υπό εξέταση μαθητών με αυτήν των υπόλοιπων συμμαθητών τους στη γλώσσα και τα μαθηματικά. Σε σχετική έρευνα, εκπαιδευτικοί χρησιμοποιώντας την κλίμακα TRAP αναγνώρισαν με ακρίβεια επιτυχώς 96,2% των περιπτώσεων που τους δόθηκαν ως μαθητές με ΜΔ και 85,7% ως μαθητές με κανονική επίδοση. Σε μια άλλη προσπάθεια ελέγχου της ικανότητας των εκπαιδευτικών να αναγνωρίσουν με ακρίβεια και επιτυχία περιπτώσεις μαθητών με ΜΔ, μαθητών με χαμηλή επίδοση στη γλώσσα και μαθητών που πιθανόν να παρουσίαζαν ΜΔ, τα αποτελέσματα έδειξαν ότι οι εκπαιδευτικοί της συγκεκριμένης μελέτης ήταν επιτυχείς στο 95% των περιπτώσεων των μαθητών αυτών (Gresham, MacMillan & Bocian, 1997).

Μια ευρύτερη εφαρμογή της αξιολογικής κρίσης των εκπαιδευτικών είναι και η αξιολογική κλίμακα Μαθησιακών Δυσκολιών (LDES, Learning Disability Evaluation Scale) από τον McCaerney (1996). Η κλίμακα περιέχει 88 ερωτήματα που έχουν χωριστεί σε επτά (7) ενότητες και αξιολογούνται από 1=ποτέ ή σπάνια, έως 3=σχεδόν πάντα. Έρευνες έχουν δείξει ότι η κλίμακα έχει επαρκή αξιοπιστία και εσωτερική στα-

θερότητα, όπως και εγκυρότητα κριτηρίου. Όπως σημειώνεται, η κλίμακα μπορεί να λειτουργήσει όχι μόνο στον εντοπισμό των μαθητών με ΜΔ αλλά και στην περιοδική επαναξιολόγηση των στόχων του παρεμβατικού προγράμματος που έχει καταρτιστεί για το μαθητή με ΜΔ (Mathew, 2001).

Εκτός από τις ΗΠΑ, η χρήση των αξιολογικών κρίσεων των εκπαιδευτικών με τη βοήθεια κλιμάκων έχει εφαρμοστεί και σε άλλα εκπαιδευτικά συστήματα. Σε διαχρονική έρευνα στη Νορβηγία, εκπαιδευτικοί αξιολόγησαν τις επιδόσεις 603 μαθητών στην ανάγνωση και την ορθογραφία. Οι κρίσεις των εκπαιδευτικών είχαν στατιστικώς σημαντική συνάφεια με τις επιδόσεις των μαθητών αυτών σε σταθμισμένα τεστ ανάγνωσης, ορθογραφίας, όπως και τεστ νοημοσύνης (Salvesen & Undheim, 1994). Επίσης, στην ίδια έρευνα, οι εκπαιδευτικοί που συμμετείχαν μπόρεσαν να αναγνωρίσουν με ακρίβεια σχεδόν το 90% των περιπτώσεων μαθητών που είχαν αναγνωστικές δυσκολίες και να τους διαφοροποιήσουν από τους συμμαθητές τους που είχαν χαμηλή επίδοση στην ανάγνωση. Σε άλλη έρευνα που διεξήχθη στην Αυστραλία, οι Kenny και Chekaluk (1993) επιχείρησαν να συγκρίνουν τις αξιολογικές κρίσεις εκπαιδευτικών (όπως αυτές είχαν προκύψει από ερωτηματολόγια και αξιολογικές κλίμακες) με τις επιδόσεις των μαθητών σε διάφορα σταθμισμένα τεστ γλώσσας και ανάγνωσης σε δείγμα μαθητών νηπιαγωγείου και των δύο πρώτων τάξεων του Δημοτικού. Στην έρευνα αυτή οι μαθητές ολοκλήρωσαν δοκιμασίες όπως Word Attack και Word Identification του τεστ Woodcock-Johnson Psychoeducational Battery (Woodcock & Johnson, 1989), το τεστ Lindamood Auditory Conceptualization, LAC (Lindamood & Lindamood, 1979) και την αναθεωρημένη έκδοση του Peabody Picture Vocabulary Test, PPVT-R (Dunn & Dunn, 1981). Τα αποτελέσματα της έρευνας αυτής έδειξαν ότι για τους μαθητές της πρώτης και της δεύτερης τάξης, οι κρίσεις των εκπαιδευτικών και ένα απλό σύστημα αξιολόγησης της αναγνωστικής δεξιότητας των μαθητών 3 επιπέδων (χαμηλή, μέτρια, υψηλή) μπορούσαν να προ-

βλέψουν την επίδοση των μαθητών αυτών στην ανάγνωση (όπως μετρήθηκε με τη βοήθεια του τεστ Woodcock-Johnson) καλύτερα από ό,τι τα άλλα δύο τεστ (δηλαδή το LAC και το PPVT-R) που χρησιμοποιήθηκαν στην έρευνα (Kenney & Chekaluk, 1993).

Πιο πρόσφατα, σε παραπλήσια έρευνα μελετήθηκε η προγνωστική εγκυρότητα των αξιολογικών κρίσεων εκπαιδευτικών του νηπιαγωγείου (Teisl, Mazocco & Myers, 2001). Οι νηπιαγωγοί του δείγματος αξιολόγησαν τις επιδόσεις των μαθητών τους σε μαθηματικές έννοιες και προαναγνωστικές δραστηριότητες. Στη συνέχεια, έγινε συσχέτιση των κρίσεων αυτών με τις επιδόσεις των μαθητών στην πρώτη τάξη. Η ανάλυση των σχετικών δεδομένων ανέδειξε υψηλές, στατιστικώς σημαντικές συνάφειες μεταξύ εκπαιδευτικών κρίσεων και επιδόσεων των μαθητών στην πρώτη τάξη. Παράλληλα, οι εκπαιδευτικοί μπόρεσαν να διακρίνουν τους μαθητές εκείνους που πιθανόν να αντιμετωπίζουν μαθησιακές δυσκολίες στην πρώτη τάξη, αν και υπήρξαν αρκετές περιπτώσεις λανθασμένων κρίσεων. Λόγω του γεγονότος αυτού, οι ερευνητές υποστήριξαν την υπό όρους και προσεκτική χρήση των αξιολογικών κρίσεων των εκπαιδευτικών παράλληλα με την εφαρμογή και άλλων τεστ γνωστικών δεξιοτήτων στους μαθητές.

Παρόμοιες μελέτες που σχετίζονταν με την πρόοδο και τη μετάβαση από το νηπιαγωγείο στις πρώτες τάξεις του Δημοτικού σχολείου και βασίστηκαν στις αξιολογικές κρίσεις των εκπαιδευτικών έγιναν από την ομάδα των Taylor, Anselmo, Foreman, Schatschneider και Angelopoulos (2000) και από τους Crosby και French (2002). Στην πρώτη έρευνα των Taylor και συνεργατών (2000), όσοι μαθητές του νηπιαγωγείου κρίθηκαν από τους εκπαιδευτικούς ότι δεν είχαν την ίδια ακαδημαϊκή πρόοδο, είχαν χαμηλότερη επίδοση σε τεστ ανάγνωσης, ορθογραφίας, μαθηματικών και άλλων δεξιοτήτων από ό,τι οι συμμαθητές τους χωρίς μαθησιακά προβλήματα. Οι διαφορές αυτές ήταν εμφανείς ακόμη και όταν οι μαθητές επαναξιολογήθηκαν προς το τέλος της πρώτης τάξης στο Δημοτικό. Ομοίως, στη μελέτη των

Crosby και French (2002), εκπαιδευτικοί αξιολόγησαν τις μαθησιακές συμπεριφορές των μαθητών από οικογένειες χαμηλού οικονομικού επιπέδου που συμμετείχαν σε πρόγραμμα ενίσχυσης από το νηπιαγωγείο μέχρι και την τρίτη τάξη του Δημοτικού. Τα αποτελέσματα έδειξαν πως υπήρχε υψηλή αξιοπιστία μεταξύ των επαναλαμβανόμενων μετρήσεων και παρά πολύ υψηλή εσωτερική συνέπεια στις κρίσεις των εκπαιδευτικών και στη δυνατότητά τους να εντοπίζουν έγκαιρα τους μαθητές εκείνους που πιθανόν να παρουσιάσουν αργότερα μαθησιακά προβλήματα.

Παρά τον ικανοποιητικό αριθμό ερευνών που αναδεικνύουν την ικανότητα των εκπαιδευτικών να λειτουργήσουν ως κριτήρια αξιολόγησης (teachers as tests), εξακολουθεί να επικρατεί μια δυσπιστία ως προς την ακρίβεια και την αξιοπιστία τους. Γενικά, οι εκπαιδευτικοί θεωρούνται κακοί κριτές των ικανοτήτων των μαθητών (Hoge & Coladarsi, 1989). Ίσως, το καλύτερο επιχείρημα για την εφαρμογή και χρήση αξιολογικών κρίσεων από εκπαιδευτικούς να προέρχεται από τους Gresham και συνεργάτες, που υποστήριξαν ότι «κάπου στην ανάπτυξη των διαδικασιών λήψης ψυχοπαιδαγωγικών αποφάσεων, οι κρίσεις των εκπαιδευτικών υποβαθμίστηκαν σε εικασίες, υπόνοιες και, πάνω από όλα, ύποπτα δεδομένα» (Gresham, MacMillan & Bocian 1997, σελ. 552). Ωστόσο, τα μέχρι στιγμής διαθέσιμα ερευνητικά δεδομένα με διαφορετικά δείγματα μαθητών με ή χωρίς ΜΔ στηρίζουν την άποψη ότι οι αξιολογικές κρίσεις των εκπαιδευτικών, που θεμελιώνονται με τη βοήθεια κάποιας κλίμακας, μπορούν να χρησιμοποιηθούν για την αναγνώριση και τον έγκαιρο εντοπισμό μαθητών με ΜΔ.

Εκπαιδευτική αξιολόγηση της γραπτής έκφρασης

Όπως προαναφέρθηκε, η αξιολογική κρίση των εκπαιδευτικών έχει αξιοπιστία, συνέπεια και υψηλή εγκυρότητα κριτηρίου συγκρινόμενη με άλλα ψυχομετρικά εργαλεία και τεστ επίδοσης. Ωστόσο, με εξαίρεση την κλίμακα LDES, η οποία περιείχε μια υποενότητα για τη γραπτή έκφραση,

καμία από τις μελέτες που παρουσιάστηκαν προηγουμένως δεν ασχολήθηκαν συστηματικά με την αξιολόγηση των γραπτών δεξιοτήτων των μαθητών. Λόγου χάρη, η κλίμακα TRAP (Gresham, Reschly & Carey, 1987) δεν εστίασε καθόλου στην παραγωγή γραπτού λόγου. Ομοίως, η κλίμακα ARPS περιείχε ερωτήματα που αφορούσαν μόνο στην αναγνωσιμότητα του γραφικού χαρακτήρα των μαθητών (DuPaul, Rapport & Perriello 1991). Αυτή η απουσία αξιολογικών κριτηρίων πιθανόν να οφείλεται στην πολυπλοκότητα της γνωστικής διαδικασίας παραγωγής γραπτού λόγου και των απαιτήσεών της (Berninger, Mizokawa & Bragg, 1991. Hayes & Flower, 1980. Bereiter & Scardamalia, 1987). Όπως τόνισε και η Weiner (1980), «...η πολυπλοκότητα της γραπτής έκφρασης καθιστά την αξιολόγηση του γραπτού κειμένου δύσκολη» (σελ. 48).

Κατά τη δεκαετία του 1980, πολλές μελέτες ασχολήθηκαν με την ανάπτυξη της γραπτής έκφρασης, την παραγωγή γραπτών κειμένων από μαθητές με ΜΔ και τους συμμαθητές τους με κανονική επίδοση, όπως και με τη συστηματική ανάλυση της γνωστικής διαδικασίας της γραπτής έκφρασης. Τις έρευνες αυτές έχουν συνοψίσει σε δύο ανεξάρτητες επισκοπήσεις της βιβλιογραφίας οι Lynch και Jones (1989) και οι Newcomer και Barenbaum (1991). Για τη σύγκριση της ικανότητας γραπτής έκφρασης των μαθητών και τη μελέτη της ανάπτυξης της παραγωγής γραπτού λόγου χρησιμοποιήθηκαν πολλά και διαφορετικά κριτήρια όπως συνολικός αριθμός λέξεων, αριθμός ορθογραφημένων λέξεων, αριθμός μονάδων ιδεών (idea units), αριθμός θεματικών μονάδων (T-units), αριθμός ρημάτων, αριθμός λέξεων ανά πρόταση κ.λπ. (Houck & Billingsley, 1989). Επίσης, χρησιμοποιήθηκαν και άλλα μη ποσοτικά και περισσότερο ποιοτικά κριτήρια, όπως η ποιότητα του κειμένου (πολύ χαμηλή έως πολύ υψηλή), η αναγνωσιμότητα του γραφικού χαρακτήρα του μαθητή, το είδος του κειμένου (πλήρες κείμενο, ελλιπές κείμενο, απλή αλληλουχία γεγονότων, μη σχετικό με το θέμα κείμενο).

Εκτός από τη χρήση εργαστηριακών κριτηρίων (όπως τα προαναφερθέντα), για την αξιολόγηση των γραπτών δεξιοτήτων των μαθητών

(και ιδίως αυτών με ΜΔ στη γλώσσα) αναπτύχθηκαν αρκετά σταθμισμένα τεστ. Μόνο το τεστ TOWL (Test of Written Language) ήταν ένα πλήρες τεστ γραπτής έκφρασης (Hammill & Larsen, 1996). Αντίθετα, άλλα τεστ ασχολούνταν μόνο με την ορθογραφημένη γραφή μεμονωμένων λέξεων, όπως το Kaufman Test of Educational Achievement-Revised (Kaufman & Kaufman, 1985) και το Wide Range Achievement Test-Revised (Jastak & Wilkinson, 1984) ή με μικρές δοκιμασίες γραφής (όπως το Woodcock Johnson). Αν και αυτά τα σταθμισμένα τεστ έχουν επαρκείς ψυχομετρικές ιδιότητες, εντούτοις δεν έχουν άμεση σχέση με το αναλυτικό πρόγραμμα που ακολουθεί ο μαθητής στην τάξη, ενώ ορισμένα από αυτά είναι περιορισμένης έκτασης και αφορούν συγκεκριμένες πτυχές της γραπτής έκφρασης. Παράλληλα, όπως όλα τα σταθμισμένα τεστ δίνουν μια συγκεκριμένη εικόνα των ικανοτήτων του μαθητή (τη συγκεκριμένη στιγμή της αξιολόγησης) και φυσικά δεν εμπλέκουν στη διαδικασία αξιολόγησης τον εκπαιδευτικό του εν λόγω μαθητή που θα μπορούσε να συνεισφέρει και ο ίδιος τη δική του αξιολογική κρίση για τις γραπτές δεξιότητες και τα προβλήματα που αντιμετωπίζει ο μαθητής.

Σε μια προσπάθεια να ξεπεραστούν τόσο η συχνά περιορισμένη έκταση των διαθέσιμων σταθμισμένων τεστ όσο και η έλλειψη σύνδεσης των τεστ με τη μαθησιακή διαδικασία και το πρόγραμμα του μαθητή, νέες, εναλλακτικές μέθοδοι αξιολόγησης έγιναν δημοφιλείς κατά τη διάρκεια της δεκαετίας του 1980 και μετά. Μια τέτοια μέθοδος ήταν και η αξιολόγηση που βασίζεται στο αναλυτικό πρόγραμμα (Curriculum Based Assessment). Με τη μέθοδο αυτή, οι εκπαιδευτικοί μπορούσαν σε τακτά χρονικά διαστήματα να αξιολογούν την επίδοση του μαθητή σε διάφορα σχολικά αντικείμενα (όπως, ανάγνωση, ορθογραφία, γραπτή έκφραση, μαθηματικά) δίνοντας μικρά και σύντομα τεστ δεξιοτήτων προσαρμοσμένα από το αναλυτικό πρόγραμμα του μαθητή (Shinn, 1989). Αυτή η επαναλαμβανόμενη διαδικασία είναι σύντομη και διαρκεί από ένα

(1) έως και τρία (3) λεπτά για κάθε αξιολογούμενο στοιχείο. επομένως είναι εφικτή η συνεχής αξιολόγηση του μαθητή και η σύγκριση της επίδοσής του με την επίδοση του ίδιου σε προηγούμενες κρίσεις ή με την επίδοση των συμμαθητών του κάθε φορά. Όσον αφορά την αξιολόγηση της γραπτής έκφρασης, με τη μέθοδο αυτή ο μαθητής γράφει σύντομα κείμενα τα οποία αξιολογούνται κυρίως με τρία κριτήρια: συνολικός αριθμός λέξεων που έγραψε, συνολικός αριθμός ορθογραφημένων λέξεων και αριθμός ορθών ακολουθιών λέξεων. Οι ψυχομετρικές ιδιότητες της μεθόδου είναι επαρκείς με τους δείκτες αξιοπιστίας να κυμαίνονται από 0,83 έως 0,94 και τους δείκτες εγκυρότητας κριτηρίου από 0,80 έως 0,99 (Marston, 1989). Παρά τις ικανοποιητικές ψυχομετρικές ιδιότητες, αυτή η μέθοδος έχει και περιορισμούς καθώς, λόγω των κριτηρίων που χρησιμοποιούνται, ο εκπαιδευτικός δεν μπορεί να αξιολογήσει την επίδοση του μαθητή στην εφαρμογή γραμματικών κανόνων ή τη θεματική ανάπτυξη, τη συνοχή και τη συνεκτικότητα του κειμένου που γράφει. Από την εκπαιδευτική κοινότητα φαίνεται να γίνεται αποδεκτή αυτή η μέθοδος αξιολόγησης του μαθητή. Σε μελέτη τους, οι Eckert, Shapiro και Lutz έδειξαν πως είναι σημαντική η σωστή κατάρτιση στη χρήση της μεθόδου αυτής για την αποδοχή της από τους εκπαιδευτικούς (1995).

Συνοψίζοντας, οι κρίσεις των εκπαιδευτικών φαίνεται να αποτελούν μια επαρκή μέθοδο αξιολόγησης των ικανοτήτων των μαθητών και να προσφέρουν μια εναλλακτική διέξοδο στα διαθέσιμα σταθμισμένα τεστ επίδοσης. Όσον αφορά το γραπτό λόγο, η χρήση σταθμισμένων τεστ έχει συγκεκριμένα πλεονεκτήματα αλλά και μειονεκτήματα. Ομοίως, πλεονεκτήματα και μειονεκτήματα φαίνεται να έχει και η μέθοδος δυναμικής αξιολόγησης που βασίζεται στο αναλυτικό πρόγραμμα (Curriculum-Based Measurement-CBM). Μέχρι στιγμής τα στοιχεία που παρουσιάστηκαν αφορούσαν την αγγλική γλώσσα (με εξαίρεση τις έρευνες που έγιναν σε χώρες όπου η αγγλική δεν είναι η πρώτη γλώσσα του μαθητή). Για την ελληνική εκπαιδευτική πραγματικό-

τητα, λίγα είναι τα διαθέσιμα στοιχεία. Σε μελέτη της Triga (2004), δείγμα Ελλήνων εκπαιδευτικών αξιολόγησε τις αναγνωστικές δεξιότητες μαθητών και οι κρίσεις αυτές συσχετίστηκαν με τις επιδόσεις των μαθητών αυτών σε σταθμισμένα τεστ ανάγνωσης. Στην έρευνα αυτή καταγράφηκαν υψηλές και στατιστικώς σημαντικές συνάφειες μεταξύ των κρίσεων των εκπαιδευτικών και της επίδοσης των μαθητών στο τεστ ανάγνωσης. Σε άλλη πρόσφατη έρευνα, που αφορούσε τη συσχέτιση κρίσεων Ελλήνων εκπαιδευτικών και την αξιολόγηση των γραπτών δεξιοτήτων των μαθητών (με τη χρήση της μεθόδου CBM), οι κρίσεις των εκπαιδευτικών στην ορθογραφία και τη γραπτή έκφραση βρέθηκαν να έχουν στατιστικώς σημαντική συνάφεια ιδίως με το κριτήριο των ορθών ακολουθιών λέξεων (Δημάκος, 2007).

Επομένως, σκοπός της μελέτης αυτής ήταν να συσχετίσει τις αξιολογικές κρίσεις των εκπαιδευτικών με τις επιδόσεις μαθητών των τελευταίων τάξεων του Δημοτικού σε σειρά κριτηρίων αξιολόγησης του γραπτού λόγου που παρουσιάστηκαν προηγουμένως και να μελετήσει την ικανότητα ή μη των εκπαιδευτικών να κρίνουν έγκυρα και αξιόπιστα τις γραπτές δεξιότητες των μαθητών τους.

2. Μέθοδος

Δείγμα

Στην έρευνα συμμετείχαν τρεις διαφορετικές ομάδες μαθητών που φοιτούσαν στις ανώτερες τάξεις του Δημοτικού. Πιο συγκεκριμένα, η πρώτη ομάδα περιλάμβανε 75 μαθητές των τριών τελευταίων τάξεων, που κατανέμονταν ως εξής: 19 μαθητές στην Δ' τάξη (10 αγόρια, 9 κορίτσια), 25 στην Ε' τάξη (9 αγόρια και 16 κορίτσια) και 31 μαθητές στην Στ' τάξη (11 αγόρια και 20 κορίτσια). Η δεύτερη ομάδα περιλάμβανε 76 μαθητές (που φοιτούσαν και αυτοί στις τρεις τελευταίες τάξεις (Δ', Ε' και Στ' τάξη) του Δημοτικού και κατανέμονταν ως εξής: 23 μαθητές στην Δ' τάξη (12 αγόρια και 11 κορίτσια), 27 μαθητές στην Ε' τάξη (14 αγόρια και 13 κορίτσια) και 26 μαθητές

στην Στ' τάξη (13 αγόρια και 13 κορίτσια). Τέλος, η τρίτη ομάδα περιλάμβανε 77 μαθητές που φοιτούσαν στην Δ' και την Στ' τάξη του Δημοτικού σχολείου και κατανέμονταν ως εξής: 40 μαθητές στην Δ' τάξη (23 αγόρια και 17 κορίτσια) και 36 μαθητές στην Στ' τάξη (18 αγόρια και 19 κορίτσια).

Όλοι οι μαθητές (και των τριών ομάδων) προέρχονταν από διαφορετικά σχολεία της Πάτρας και των όμορων περιοχών, είχαν την ελληνική ως πρώτη ή κύρια γλώσσα, δεν είχαν επαναλάβει καμία τάξη στο δημοτικό σχολείο και δεν παρουσίαζαν μαθησιακές δυσκολίες ή κάποιες άλλες ειδικές εκπαιδευτικές ανάγκες διαγνωσμένες από επίσημο κρατικό φορέα.

Το σύνολο των μαθητών είχε στο παρελθόν συμμετάσχει σε προηγούμενες έρευνες του Τομέα Ψυχολογίας σχετικά με την ανάπτυξη των δεξιοτήτων γραπτής έκφρασης των μαθητών σχολικής ηλικίας και την ικανότητα των εκπαιδευτικών τους να αξιολογούν αξιόπιστα και έγκυρα τις δεξιότητες αυτές των μαθητών τους. Στην περίπτωση του παρόντος έργου, χρησιμοποιήθηκαν τα δεδομένα εκείνα που σχετίζονταν με τη μελέτη της αξιοπιστίας και της εγκυρότητας των κρίσεων των εκπαιδευτικών.

Υλικά

1. Στους μαθητές και των τριών ομάδων δόθηκε μια ατελής πρόταση με σκοπό να τη συμπληρώσουν και να συνεχίσουν να γράψουν ένα κείμενο αφηγηματικού είδους. Σε κάθε ομάδα δόθηκε διαφορετική πρόταση-έναυσμα. Πιο συγκεκριμένα, στην πρώτη ομάδα δόθηκε η ακόλουθη πρόταση: «Την τελευταία φορά που πήγα εκδρομή...», στη δεύτερη ομάδα: «Οι ωραιότερες διακοπές που έχω κάνει ήταν...» και στην τρίτη ομάδα: «Τα παιδιά ετοιμάστηκαν να πάνε μια βόλτα...». Όπως προαναφέρθηκε, οι τρεις ομάδες μαθητών είχαν στο παρελθόν συμμετάσχει σε προηγούμενες έρευνες σχετικά με την ανάπτυξη του γραπτού λόγου στις οποίες οι μαθητές αυτοί έγραψαν γραπτά κείμενα αφηγηματικού τύπου. Έτσι, αν και οι ατελείς προτάσεις που

δόθηκαν ως έναυσματα ήταν διαφορετικές σε κάθε περίπτωση, εντούτοις αφορούσαν όλες το ίδιο κειμενικό είδος (αφήγηση) και είχαν και οι τρεις προτάσεις περίπου το ίδιο αντικείμενο (μια εκδρομή, μια βόλτα ή διακοπές).

2. Οι εκπαιδευτικοί των 8 τάξεων στις οποίες φοιτούσαν οι μαθητές του παρόντος έργου συμπλήρωσαν μία σύντομη κλίμακα τύπου Likert 5 σημείων (όπου 1=πολύ χαμηλή επίδοση και 5=πολύ υψηλή επίδοση) που αφορούσε στην επίδοση του μαθητή στη γλώσσα και ιδιαίτερα στους τομείς της ανάγνωσης, της ορθογραφικής γραφής και της γραπτής έκφρασης.

Διαδικασία

Οι μαθητές έλαβαν οδηγίες να σκεφτούν για σύντομο χρονικό διάστημα σχετικά με την ατελή πρόταση που τους είχε δοθεί και στη συνέχεια να ολοκληρώσουν την πρόταση και να γράψουν ένα μικρό σχετικό κείμενο. Ενώ οι μαθητές ολοκλήρωναν το γραπτό κείμενό τους, οι οκτώ εμπλεκόμενοι εκπαιδευτικοί συμπλήρωναν τη σχετική κλίμακα αξιολογικής κρίσης των γλωσσικών δεξιοτήτων των μαθητών των τάξεών τους. Πιο συγκεκριμένα στους εκπαιδευτικούς δόθηκαν οι εξής οδηγίες: «Αξιολογείτε τη γενικότερη επίδοση στην ορθογραφία, την ανάγνωση και τη γραπτή έκφραση του κάθε μαθητή με τη βοήθεια της ακόλουθης κλίμακας (όπου 1=πολύ χαμηλή και 5=πολύ υψηλή) με βάση την αναμενόμενη για την ηλικία και τη σχολική τάξη επίδοση». Οι οδηγίες αυτές ήταν σύμφωνες και με τις αντίστοιχες οδηγίες στην περίπτωση της κλίμακας TRAP (Gresham, Reschly & Carey 1987). Κανένας άλλος λειτουργικός ορισμός δεν δόθηκε στους εκπαιδευτικούς.

Ανάλυση

Το κείμενο της πρώτης και της τρίτης ομάδας κρίθηκε με βάση τα εξής κριτήρια: συνολικός αριθμός λέξεων, συνολικός αριθμός ορθογραφημένων λέξεων. Επίσης στο κείμενο της

Πίνακας 1
Δείκτες Cronbach (a) και συνάφειες συνόλου-κριτηρίου (item-total)
μεταξύ αξιολογικών κρίσεων εκπαιδευτικών

Ομάδα	Cronbach's a	Ανάγνωση	Ορθογραφία	Γραπτή έκφραση
N=75	0,9307	0,8924	0,8902	0,7955
N=76	0,8969	0,7490	0,7873	0,8597
N=77	0,8690	0,8303	0,6610	0,7634

πρώτης ομάδας εφαρμόστηκαν και τα εξής κριτήρια: συνολικός αριθμός μονάδων ιδεών, ποιότητα του κειμένου με τη χρήση κλίμακας Likert 5 σημείων (όπου 1 = πολύ χαμηλή ποιότητα και 5 = πολύ υψηλή ποιότητα). Για το κείμενο της δεύτερης ομάδας χρησιμοποιήθηκαν κριτήρια της μεθόδου CBM και πιο συγκεκριμένα, συνολικός αριθμός λέξεων, συνολικός αριθμός ορθογραφημένων λέξεων, αριθμός ορθών ακολουθιών λέξεων. Επίσης, η ανάλυση της κανονικότητας των κατανομών των μεταβλητών της παρούσας έρευνας έδειξε ότι οι μεταβλητές της εργασίας ακολουθούσαν σε γενικές γραμμές την κανονική κατανομή με εξαίρεση τη μεταβλητή Αριθμός Ορθογραφικών Λαθών (στην πρώτη και τρίτη ομάδα) για την οποία το κριτήριο Shapiro-Wilk έδειξε έλλειψη κανονικότητας ($w=0,7369$ και $w=0,5976$, αντίστοιχα).

3. Αποτελέσματα

Ο πίνακας 1 παρουσιάζει βασικά ψυχομετρικά χαρακτηριστικά για τις αξιολογικές κρίσεις των εκπαιδευτικών ξεχωριστά για κάθε ομάδα.

Όπως φαίνεται και από την ανάγνωση των στοιχείων του πίνακα, και για τις τρεις ομάδες μαθητών, οι δείκτες α του Cronbach είναι ικανοποιητικοί, ενώ πολύ υψηλοί είναι και οι δείκτες συνάφειας κριτηρίου-συνόλου (item-total) για τα 3 αξιολογικά κριτήρια που ερευνώνται. Στον πίνακα 2 παρουσιάζονται οι μεταξύ των κριτηρίων συνάφειες που και αυτές είναι στατιστικώς σημαντικές ($p < 0,0005$) σε κάθε περίπτωση.

Τα παραπάνω αποτελέσματα ενισχύουν την εικόνα της σταθερότητας των αξιολογικών κρίσεων των εκπαιδευτικών και για τις τρεις ομάδες, συνολικά. Ωστόσο, λόγω των υψηλών και στατιστικά σημαντικών συναφειών (που κυμαίνονταν από 0,5856 μεταξύ ορθογραφίας και γραπτής έκφρασης έως 0,9056 μεταξύ ορθογραφίας και ανάγνωσης) των κρίσεων των εκπαιδευτικών στους τρεις τομείς της γλώσσας (ανάγνωση, ορθογραφία, γραπτή έκφραση), δεν είναι απολύτως σωστή η χρήση αυτών των συναφειών στον έλεγχο της εγκυρότητας κριτηρίου των αξιολογικών κρίσεων. Έτσι στην επόμενη φάση ακολουθήθηκε η εξής διαδικασία.

Για κάθε ζεύγος αξιολογικής κρίσης εκπαιδευτικών και κριτηρίου αξιολόγησης του κειμένου, υπολογίστηκε ο δείκτης μερικής συνάφειας μεταξύ των δύο αυτών στοιχείων, συγκεκριμένα χρησιμοποιήθηκαν οι δείκτες μερικής συνάφειας (partial correlation) μεταξύ των κρίσεων των εκπαιδευτικών και των διαφόρων κριτηρίων αξιολόγησης των γραπτών κειμένων που έχουν κατά καιρούς προταθεί, ελέγχοντας κάθε φορά την επίδραση των αξιολογικών κρίσεων στους άλλους δύο τομείς της γλώσσας. Αυτοί οι δείκτες μερικής συνάφειας παρουσιάζονται στον πίνακα 3 που ακολουθεί. Όπως φαίνεται, δεν υπάρχει καμία στατιστικώς σημαντική συνάφεια μεταξύ οποιουδήποτε αξιολογικού κριτηρίου των κειμένων και της κρίσης των εκπαιδευτικών στην ανάγνωση ($p > 0,05$).

Επίσης, η κρίση των εκπαιδευτικών για την ικανότητα ορθογραφημένης γραφής των μαθητών συσχετίζεται μερικώς μόνο με τον αριθμό των ορθογραφικών λαθών και στις τρεις ομάδες

Πίνακας 2
Βασικά περιγραφικά στατιστικά και συντελεστές
συνάφειας Pearson μεταξύ κρίσεων εκπαιδευτικών

Κρίσεις εκπαιδευτικών	M (sd)	Ορθογραφία	Γραπτή έκφραση
Πρώτη Ομάδα (N=75)			
Ανάγνωση	3.68 (1.14)	0.9056	0.7763
Ορθογραφία	3.48 (1.22)	–	0.7767
Γραπτή έκφραση	3.12 (1.21)	–	–
Δεύτερη Ομάδα (N=76)			
Ανάγνωση	4.18 (0.99)	0.6639	0.7596
Ορθογραφία	3.79 (1.07)	–	0.8069
Γραπτή έκφραση	4.08 (1.08)	–	–
Τρίτη Ομάδα (N=77)			
Ανάγνωση	3.75 (1.14)	0.6706	0.8083
Ορθογραφία	3.64 (1.12)	–	0.5856
Γραπτή έκφραση	3.31 (1.12)	–	–

Πίνακας 3
Συντελεστές μερικής συνάφειας μεταξύ κρίσεων εκπαιδευτικών
και κριτηρίων αξιολόγησης κειμένων

Κριτήρια κειμένων	Αξιολογικές κρίσεις εκπαιδευτικών		
	Ανάγνωση	Ορθογραφία	Γραπτή έκφραση
Πρώτη Ομάδα (N=75)			
Συν. αριθμός λέξεων	–	–	0.4407 ($p < 0.0001$)
Αριθμός ορθογρ. λαθών	–	-0.3614 ($p < 0.001$)	–
Αριθμός ορθογραφημένων λέξεων	–	–	0.4288 ($p < 0.0001$)
Συν. μονάδων ιδεών	–	–	0.3583 ($p < 0.002$)
Δεύτερη ομάδα (N=76)			
Συν. αριθμός λέξεων	–	–	–
Αριθμός ορθογραφημένων λέξεων	–	0.2551 ($p < 0.026$)	–
Ορθές ακολουθίες λέξεων	–	–	0.3046 ($p < 0.008$)
Τρίτη ομάδα (N=77)			
Συν. αριθμός λέξεων	–	0.4408 ($p < 0.0001$)	0.2743 ($p < 0.016$)
Αριθμός ορθογρ. λαθών	–	-0.4422 ($p < 0.0001$)	–
Αριθμός ορθογραφημένων λέξεων	–	0.3853 ($p < 0.0005$)	0.2777 ($p < 0.015$)

της έρευνας ($r=0,3614$, $r=0,2551$ και $r=0,4422$, για καθεμιά ομάδα αντίστοιχα). Επιπροσθέτως, στην περίπτωση της τρίτης ομάδας συσχετίζεται μερικώς και με τα υπόλοιπα αξιολογικά κριτήρια του κειμένου (συνολικός αριθμός λέξεων: $r=0,4408$, και αριθμός ορθογραφημένων λέξεων: $r=0,3853$, αντίστοιχα).

Όσον αφορά την κρίση των εκπαιδευτικών για την ικανότητα γραπτής έκφρασης, οι δείκτες μερικής συνάφειας ήταν γενικά μεσαίου μεγέθους επίδρασης (effect size), αλλά στατιστικώς σημαντικοί, και κυμάνθηκαν από $r=0,2743$ για το συνολικό αριθμό λέξεων του κειμένου της τρίτης ομάδας μέχρι $r=0,4407$ για το ίδιο κριτήριο στην πρώτη ομάδα του δείγματος.

4. Συμπεράσματα

Συνοψίζοντας τα προαναφερθέντα αποτελέσματα της παρούσας έρευνας που βασίστηκε σε τρεις διαφορετικές ομάδες μαθητών του Δημοτικού, παρατηρούνται τα εξής. Πρώτον, οι κρίσεις των εκπαιδευτικών παρουσίασαν πολύ υψηλούς συντελεστές εσωτερικής συνέπειας *a* του Cronbach σε κάθε περίπτωση και για όλες τις ομάδες. Δεύτερον, εξίσου υψηλοί ήταν και οι συντελεστές συνάφειας συνόλου-κριτηρίου (item-total) για τις κρίσεις των εκπαιδευτικών και στις τρεις ομάδες μαθητών. Τρίτον, τα δεδομένα αυτά υποστηρίζονται και από εξίσου υψηλούς συντελεστές συνάφειας μεταξύ των κρίσεων των εκπαιδευτικών στην ανάγνωση, την ορθογραφημένη γραφή και τη γραπτή έκφραση. Τέταρτον, χρησιμοποιώντας συντελεστές μερικής συνάφειας (partial order correlation coefficients) μεταξύ των κρίσεων των εκπαιδευτικών και σειράς κριτηρίων για τον έλεγχο των κειμένων που έγραψαν οι μαθητές, παρατηρήθηκαν μεσαίου μεγέθους συσχετίσεις μεταξύ της κρίσης των εκπαιδευτικών για τη γραπτή έκφραση και των κριτηρίων αυτών. Πέμπτον, στο ίδιο επίπεδο κυμάνθηκαν και οι συσχετίσεις μεταξύ κρίσεων εκπαιδευτικών για την ορθογραφία και των εν λόγω κριτηρίων. Έκτον, δεν βρέθηκαν στατιστικώς ση-

μαντικοί συντελεστές συνάφειας μεταξύ των κρίσεων των εκπαιδευτικών για την ανάγνωση και των σχετικών κριτηρίων των γραπτών κειμένων.

Τα αποτελέσματα αυτά έρχονται να προστεθούν σε ένα μακρύ κατάλογο ερευνητικών δεδομένων (ενδεικτικά, DuPaul, Rapport & Perriello 1991. Gresham, Reschly & Carey 1987, 1997. Kenny & Chekaluk, 1993. Salvesen & Undheim, 1994) που έχουν ήδη αναδείξει την επάρκεια των αξιολογικών κρίσεων των εκπαιδευτικών και τη δυνατότητα εφαρμογής αυτών των κρίσεων για λόγους διαγνωστικής αξιολόγησης σε άλλα εκπαιδευτικά συστήματα. Αν και στην παρούσα μελέτη δεν κατέστη εφικτός ο υπολογισμός της αξιοπιστίας επαναλαμβανόμενων μετρήσεων (test-retest reliability) των κρίσεων των εκπαιδευτικών (και αυτό καταγράφεται στους περιορισμούς της μελέτης αυτής), οι υπόλοιποι δείκτες εσωτερικής συνέπειας (internal consistency) υποδηλώνουν πως, μάλλον, οι αξιολογικές κρίσεις των εκπαιδευτικών είναι εν μέρει αξιόπιστες ως εργαλεία για την αξιολόγηση των ικανοτήτων των μαθητών και θα μπορούσαν να αποτελέσουν τμήμα μιας συνολικότερης διαγνωστικής αξιολόγησης των μαθησιακών δυσκολιών αλλά και δυνατοτήτων των μαθητών. Ωστόσο, μελλοντικές έρευνες αξιοπιστίας επαναλαμβανόμενων μετρήσεων θα προσέφεραν περισσότερες πληροφορίες όσον αφορά τη δυνατότητα εφαρμογής και χρήσης των κρίσεων των εκπαιδευτικών.

Είναι ενθαρρυντικό το γεγονός πως οι μερικοί συντελεστές συνάφειας ήταν στατιστικά σημαντικοί μόνο μεταξύ των κριτηρίων γραπτής έκφρασης και των αξιολογικών κρίσεων των εκπαιδευτικών σχετικά με την ορθογραφία και τη γραπτή έκφραση (και όχι την ανάγνωση). Αν, λόγου χάρη, στις αξιολογικές κρίσεις συμπεριλαμβανόταν και η αξιολογική εκτίμηση για την επίδοση των μαθητών στα μαθηματικά, θα έπρεπε ο μερικός δείκτης συνάφειας μεταξύ αξιολογικής κρίσης στα μαθηματικά και των κριτηρίων των κειμένων να είναι στατιστικά μη σημαντικός και αυτός. Κάτι τέτοιο δείχνει ότι πιθανόν οι αξιολογικές κρίσεις των εκπαιδευτικών να σχετίζονται με συγκεκριμένα στοιχεία της γραπτής έκφρασης

και να διαφοροποιούνται από τη γενικότερη σχολική επίδοση και εικόνα που έχει σχηματίσει ο εκπαιδευτικός για τον κάθε μαθητή της τάξης του.

Όσον αφορά τον έλεγχο της εγκυρότητας κριτηρίου για τις αξιολογικές κρίσεις των εκπαιδευτικών, αυτός επιτεύχθηκε με τη χρήση μια σειράς κριτηρίων (συνολικός αριθμός γραπτών λέξεων, αριθμός ορθογραφικών λαθών και μονάδες ιδεών) που είχαν ήδη εφαρμοστεί από τη δεκαετία του 1980 στις τότε μελέτες της γραπτής έκφρασης και παρουσιάστηκαν στις έρευνες των Houck και Billingsley (1989), Lynch και Jones (1989) και Newcomer και Barenbaum (1991). Επίσης, χρησιμοποιήθηκαν και κριτήρια από τον τομέα της δυναμικής αξιολόγησης και πιο συγκεκριμένα από τη μέθοδο CBM. Τα αποτελέσματα έδειξαν ότι υπήρξε ικανοποιητική εγκυρότητα κριτηρίου με βάση τους συντελεστές μερικής συνάφειας μεταξύ των κρίσεων των εκπαιδευτικών και των αξιολογικών κριτηρίων των κειμένων. Η επιλογή συντελεστών μερικής συνάφειας οδήγησε σε συσχετίσεις μικρότερου μεγέθους από ό,τι οι συνάφειες που καταγράφηκαν σε προηγούμενη μελέτη (Δημάκος, 2007). Σε γενικές γραμμές, το μέγεθος επίδρασης για τις αξιολογικές κρίσεις των εκπαιδευτικών ήταν μικρού έως μεσαίου μεγέθους (όπως φαίνεται και στον πίνακα 3 του παρόντος). Πιθανόν, αυτό να οφείλεται στις συγκεκριμένες οδηγίες που δόθηκαν στους εκπαιδευτικούς να κρίνουν τους μαθητές στην ανάγνωση, τη γραπτή έκφραση και την ορθογραφία σύμφωνα με την αναμενόμενη για την ηλικία και σχολική τάξη επίδοση. Καθώς δεν υπήρξαν άλλες διευκρινιστικές οδηγίες, ενδέχεται οι εκπαιδευτικοί να ήταν σχετικά αόριστοι και λιγότερο συγκεκριμένοι στις κρίσεις τους. Μελλοντικές έρευνες με αναλυτικότερες και πιο συγκεκριμένες οδηγίες για την υπό αξιολόγηση συμπεριφορά (π.χ. συγκεκριμένος αριθμός λαθών, λέξεων, έκταση του κειμένου, εφαρμογή κανόνων γραμματικής και συντακτικού) θα διαφωτίσουν ως προς την ικανότητα των εκπαιδευτικών να αξιολογούν με αξιοπιστία και εγκυρότητα τους μαθητές τους.

Ωστόσο, πρέπει να επισημανθούν ορισμένοι περιορισμοί στο παρόν μεθοδολογικό σχέδιο. Πρώτον,

δεν υπολογίστηκε η αξιοπιστία επαναλαμβανόμενων μετρήσεων για τις κρίσεις των εκπαιδευτικών. Αυτό είναι ένα πρόβλημα γιατί λόγω του έργου του και της συνεχούς επαφής με το μαθητή ο εκπαιδευτικός πιθανόν να έχει άλλες απαιτήσεις και να περιμένει ένα διαφορετικό επίπεδο ικανότητας του μαθητή με την πάροδο του χρόνου. Με άλλα λόγια, ο εκπαιδευτικός θα κρίνει διαφορετικά το μαθητή στις αρχές του σχολικού έτους και διαφορετικά, ίσως, έναν ή και περισσότερους μήνες αργότερα. Μεταβάλλοντας το χρονικό διάστημα που παρεμβάλλεται μεταξύ των δύο μετρήσεων ίσως να βοηθήσει στη σωστότερη μελέτη της αξιοπιστίας των κρίσεων των εκπαιδευτικών. Εντούτοις, αξίζει να σημειωθεί ότι ούτε στις μελέτες της κλίμακας TRAP (Gresham, Reschly & Carey 1987, Gresham, MacMillan & Bocian 1997) αναφέρονταν δείκτες αξιοπιστίας επαναλαμβανόμενων μετρήσεων. Μόνο η έρευνα των DuPaul, Rapport & Perriello (DuPaul, Rapport & Perriello, 1991) με την κλίμακα ARPS ανέφερε αξιοπιστία επαναλαμβανόμενων μετρήσεων, ωστόσο η κλίμακα αυτή δεν εστίασε συστηματικά στη γραπτή έκφραση.

Ένα άλλο στοιχείο αναφερόταν στην κατανομή των υποκειμένων σε ηλικιακές ομάδες και κατά πόσο αυτή επηρεάζει τα εξαγόμενα αποτελέσματα. Λόγω του περιορισμένου μεγέθους και των τριών ομάδων, δεν έγινε ανάλυση ξεχωριστά ανά τάξη. Επόμενες μελέτες με μεγαλύτερο δείγμα μαθητών θα μπορέσουν να απαντήσουν στο ερώτημα αυτό. Τρίτον, δεν υπήρχε ένα σταθμισμένο κριτήριο για την αξιολόγηση της γραπτής έκφρασης που να μπορέσει να ελέγξει περαιτέρω την εγκυρότητα των κρίσεων των εκπαιδευτικών. Πρόσφατα, το Πανεπιστήμιο Πατρών ολοκλήρωσε την κατασκευή σχετικού σταθμισμένου κριτηρίου για τη διαγνωστική αξιολόγηση-διερεύνηση των μαθησιακών δυσκολιών στη γραπτή έκφραση μαθητών που φοιτούν στις τάξεις Γ΄ έως και ΣΤ΄ του Δημοτικού στα πλαίσια έργου ΕΠΕΑΕΚ (Πόρποδας κ.ά., 2008). Το εργαλείο αυτό θα μπορούσε ενδεχομένως να χρησιμοποιηθεί για την περαιτέρω μελέτη της εγκυρότητας των κρίσεων των εκπαιδευτικών για τις ικανότητες γραπτής έκφρασης των μαθητών τους.

Εκπαιδευτικές προεκτάσεις

Τελικά, είναι δυνατή η χρήση και η εφαρμογή των αξιολογικών κρίσεων των εκπαιδευτικών στην περίπτωση της γραπτής γλώσσας των μαθητών τους; Αυτή η γνώση θα ήταν χρήσιμη και σημαντική γιατί θα επέτρεπε την ουσιαστικότερη συμμετοχή των εκπαιδευτικών στη μαθησιακή αξιολόγηση των μαθητών, ιδίως αυτών που εμφανίζουν μαθησιακές δυσκολίες και προβλήματα. Ο εκπαιδευτικός έρχεται σε καθημερινή επαφή με το μαθητή και είναι σε θέση να γνωρίζει τις ικανότητες και τις αδυναμίες του μαθητή. Τα δεδομένα που παρουσιάστηκαν δείχνουν ότι μάλλον ο εκπαιδευτικός μπορεί να κρίνει εν μέρει την ικανότητα γραπτής έκφρασης των μαθητών του. Ωστόσο, απαιτούνται περισσότερες μελέτες που θα συσχετίσουν τις αξιολογικές κρίσεις των εκπαιδευτικών και με άλλα πιο ουσιαστικά κριτήρια γραπτών κειμένων (συνοχή, συνεκτικότητα κ.ά.).

Όσον αφορά το ελληνικό εκπαιδευτικό σύστημα, η δυνατότητα χρήσης των αξιολογικών κρίσεων των εκπαιδευτικών απαιτεί περαιτέρω διερεύνηση και στατιστική τεκμηρίωση αλλά και ψυχομετρική κατάρτιση των εκπαιδευτικών. Απαιτείται επίσης η μελέτη της εγκυρότητας κριτηρίου των κρίσεων με τη βοήθεια σταθμισμένου διαγνωστικού κριτηρίου γραπτής έκφρασης (που θα λειτουργήσει ως το εξωτερικό κριτήριο εγκυρότητας). Χωρίς να απορρίπτονται αλλά ούτε και να γίνονται πλήρως αποδεκτές (λόγω της έλλειψης συγκεκριμένων ψυχομετρικών δεδομένων), οι αξιολογικές κρίσεις των εκπαιδευτικών θα μπορούσαν να αποτελέσουν ένα επιπρόσθετο κριτήριο στη διεπιστημονική διαγνωστική αξιολόγηση των μαθητών με μαθησιακές δυσκολίες.

Βιβλιογραφία

- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. & Rescorla, L. A. (2002). Εγχειρίδιο για τα ερωτηματολόγια και προφίλ σχολικής ηλικίας του ΣΑΕΒΑ (μψ.). Αθήνα: Ελληνικά Γράμματα.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: LEA.
- Berninger, V. W., Mizokawa, D. T. & Bragg, R. (1991). Theory-based diagnosis and remediation of writing disabilities. *Journal of School Psychology, 29*, 57-79.
- Crosby, E. G. & French, J. L. (2002). Psychometric data for teacher judgments regarding the learning behaviors of primary grade children. *Psychology in the schools, 39*, 235-244.
- Δημάκος, Ι. Κ. (2007). Εναλλακτική αξιολόγηση των γραπτών δεξιοτήτων των μαθητών του δημοτικού. Στο Μ. Βλασσοπούλου, Α. Γιαννετοπούλου, Μ. Διαμαντή, Α. Κιρπότην, Ε. Λεβαντή, Κ. Λευθέρη & Γ. Σακελλαρίου (Επιμ.), *Γλωσσικές δυσκολίες και γραπτός λόγος στο πλαίσιο της σχολικής μάθησης* (σελ. 154-163). Αθήνα: Εκδόσεις Γρηγόρη.
- Dunn, L. M. & Dunn, L. M. (1981). *The Peabody Picture Vocabulary Test-Revised*. Pines, MN: American Guidance Service.
- DuPaul, G. J., Rapport, M. D. & Perriello, L. M. (1991). Teacher ratings of academic skills: The development of the Academic Performance Rating Scale. *School Psychology Review, 20*, 284-300.
- Eckert, T. L., Shapiro, E. S. & Lutz, J. G. (1995). Teachers' ratings of the acceptability of curriculum-based assessment methods. *School Psychology Review, 24*, 497-511.
- Gresham, F. M., MacMillan, D. L. & Bocian, K. M. (1997). Teachers as "tests": Differential ability of teacher judgments in identifying students at-risk for learning disabilities. *School Psychology Review, 26*, 47-60.
- Gresham, F. M., Reschly, D. J. & Carey, M. P. (1987). Teachers as "tests": Classification accuracy and concurrent validation in the identification of learning disabled children. *School Psychology Review, 16*, 543-553.
- Hammill, D. D. & Larsen, S. C. (1996). *Test of Written Language* (3rd edition). Austin: Pro-ed.

- Hayes, J. R. & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Sternberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hoge, R. D. (1983). Psychometric properties of teacher-judgment measures of pupil attitudes, classroom behaviors, and achievement levels. *The Journal of Special Education, 17*, 401-429.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*, 297-313.
- Houck, C. K. & Billingsley, B. S. (1989). Written expression of students with and without learning disabilities: Differences across grades. *Journal of Learning Disabilities, 22*, 561-567, 572.
- Jastak, S. & Wilkinson, G. S. (1984). *Wide Range Achievement Test-Revised*. Wilmington, DE: Jastak Associates.
- Kaufman, A. S. & Kaufman, N. L. (1985). *Kaufman Test of Educational Achievement-Revised*. Circle Pines, MN: American Guidance Service.
- Kenny, D. T. & Chekaluk, E. (1993). Early reading performance: A comparison of teacher-based and test-based assessments. *Journal of Learning Disabilities, 26*, 227-236.
- Lindamood, C. & Lindamood, P. (1979). *Lindamood Auditory Conceptualization Test*. Hingham, MA: Teaching Resources Corporation.
- Lynch, E. M. & Jones, S. D. (1989). Process and product: A review of the research on LD children's writing skills. *Learning Disability Quarterly, 12*, 74-86.
- Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it. In M. R. Shinn (Ed.), *Curriculum-Based Measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.
- Mathew, S. T. (2001). A review of the Learning Disability Evaluation Scale (LDES). *Journal of School Psychology, 39*, 279-284.
- McCarney, S. B. (1996). *The Learning Disability Evaluation Scale-renormed*. Columbia, MO: Hawthorne Educational Services.
- Newcomer, P. L. & Barenbaum, E. M. (1991). The written composing ability of children with learning disabilities: A review of the literature from 1980 to 1990. *Journal of Learning Disabilities, 24*, 578-593.
- Πόρποδας, Κ. Δ., Διακογιώργη, Κ., Δημάκος, Ι. & Καραντζής, Ι. (2008). *Εργαλείο Διαγνωστικής Διερεύνησης Δυσκολιών στη Γραπτή Έκφραση των Μαθητών Γ' - Στ' Δημοτικού*. Πάτρα: ΕΠΕΑΕΚ.
- Salvesen, K. & Undheim, J. O. (1994). Screening for learning disabilities with teacher rating scales. *Journal of Learning Disabilities, 27*, 60-66.
- Shinn, M. R. (Ed.). (1989). *Curriculum-Based Measurement: Assessing special children*. New York: Guilford Press.
- Taylor, H. G., Anselmo, M., Foreman, A. L., Schatschneider, C. & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities, 33*, 200-210.
- Teisl, J. T., Mazzocco, M. M. & Myers, G. F. (2001). The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Journal of Learning Disabilities, 34*, 286-293.
- Terman, L. M. & Childs, H. G. (1912). A tentative revision and extension of the Binet-Simon Measuring Scale of Intelligence: Part I. Introduction. *Journal of Educational Psychology, 3*, 61-74.
- Triga, A. (2004). An analysis of teachers' rating scales as sources of evidence for a standardized Greek reading test. *Journal of Research in Reading, 27*, 311-320.
- Weiner, E. S. (1980). Diagnostic evaluation of writing skills. *Journal of Learning Disabilities, 13*, 48-53.
- Woodcock, R. W. & Johnson, M. B. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised*. Allen, TX: DLM/Teaching Resources.

Assessment of written expression by teachers: reliability and validity

IOANNIS K. DIMAKOS¹

ABSTRACT

The purpose of this study is to present an alternative method to the standardized, diagnostic tools of learning abilities in writing. Internationally, the idea of teachers as tests, i.e., teachers used as evaluation criteria, is supported by several studies of various academic domains across many educational systems in different countries. In the present study, past research is briefly reviewed, followed by data from the Greek educational system. Three different sets of data are analysed. In total, 228 students attending the last 3 grades in elementary schools wrote passages on various topics which were evaluated using a number of criteria, such as total words written, correctly spelled words, correct word sequences, idea units, and so on. Following, the students' teachers evaluated the students' abilities in reading, spelling, and writing using a 5-point Likert scale. Data suggested that teacher ratings had high internal consistency (based on *Cronbach's a* coefficient) and were statistically correlated with the assessment criteria used in the evaluation of students' written texts. Although more data are needed in order to verify the findings of the present study, teachers' active and systematic involvement in the evaluation process of students' abilities and learning disabilities is feasible and may offer valuable information which may be missed during the standardized psychometric assessment of students.

Key words: Teachers, Assessment, Criteria

1. Address: Ioannis Dimakos, University of Patras, Department of Elementary Education, 26500, Patras. e-mail: idimakos@upatras.gr.