



ΚΥΠΡΙΑΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΠΟΛΙΤΙΣΜΟΥ

14 Απριλίου 2018

**Οι θέσεις της Επιστημονικής Επιτροπής για τη Στατιστική Επεξεργασία των Γραπτών
Εξετάσεων 2017 για Εγγραφή και Κατάταξη στους Πίνακες Διορισίμων**

Το Υπουργείο Παιδείας και Πολιτισμού (ΥΠΠ), ενόψει ανακριβών δημοσιευμάτων που αφορούν στη στατιστική επεξεργασία των Γραπτών Εξετάσεων 2017 για Εγγραφή και Κατάταξη στους Πίνακες Διορισίμων, ενημερώνει για τις θέσεις της Επιστημονικής Επιτροπής για τη Στατιστική Επεξεργασία των εν λόγω εξετάσεων, όπως εκφράζονται δια των εκπροσώπων της Επιτροπής, ακαδημαϊκών του Πανεπιστημίου Κύπρου, Ιάσονα Λαμπριανού και Γιώργου Σπανούδη.

Το πιο κάτω κείμενο έχει σκοπό να εξηγήσει τα χαρακτηριστικά γνωρίσματα της στατιστικής επεξεργασίας που χρησιμοποίησε η Υπηρεσία Εξετάσεων για τις Εξετάσεις Διορισίμων και να απαντήσει σε μη επιστημονικά τεκμηριωμένα δημοσιεύματα των τελευταίων ημερών.

Του κυρίως κειμένου προηγείται μια σύνοψη, η οποία παρουσιάζει τα κυριότερα σημεία, τα οποία ακολούθως επεξηγούνται πιο αναλυτικά.

1. Σύνοψη

1. *Θα μπορούσε να μη γίνει στατιστική επεξεργασία και να χρησιμοποιηθούν απλά οι αρχικές βαθμολογίες των υποψηφίων σε κάθε αντικείμενο εξέτασης;*

Κατηγορηματικά όχι. Επειδή ο σχετικός Νόμος προβλέπει τη δημιουργία καταλόγου κατάταξης των επιτυχόντων και προνοεί ότι οι επιτυχόντες μπορούν να διατηρήσουν μέχρι και για 10 έτη τη μοριοδότηση που πέτυχαν. Ως εκ τούτου, η στατιστική επεξεργασία καθίσταται αναπόφευκτη.

Η Υπηρεσία Εξετάσεων απλά εφάρμοσε τις διεθνείς πρακτικές, όπως αυτές προβλέπονται από όλους τους καταξιωμένους διεθνείς επιστημονικούς φορείς. Η στατιστική επεξεργασία των αρχικών βαθμολογιών υπαγορεύεται στον τόμο Brennan (2006) του Εθνικού Συμβουλίου Εκπαιδευτικής Μέτρησης των Ηνωμένων Πολιτειών (National Council on Measurement in Education) και του Αμερικανικού Συμβουλίου Εκπαίδευσης (American Council on Education). Επίσης υπαγορεύεται στις οδηγίες του Ευρωπαϊκού Συνδέσμου

Εκπαιδευτικής Αξιολόγησης (Association for Educational Assessment – Europe), στις οδηγίες της Διεθνούς Επιτροπής Εξετάσεων (International Test Commission), στις αναθεωρημένες οδηγίες του Ολλανδικού συστήματος Ποιότητας Εξετάσεων (Revised Dutch Rating System for Test Quality) κτλ.

2. Υπάρχουν καταξιωμένοι επιστήμονες/ερευνητές ή οργανωμένοι θεσμοί επιστημονικών ομάδων διεθνώς που να ισχυρίζονται ότι σε ανταγωνιστικές εξετάσεις κατάταξης θα πρέπει να αθροίζονται οι αρχικές – μη επεξεργασμένες – βαθμολογίες των υποψηφίων;

Κατηγορηματικά όχι. Δεν υπάρχουν επιστημονικά συγγράμματα που να προτείνουν την άθροιση αρχικών βαθμολογιών για εξετάσεις κατάταξης.

3. Μόνο η Υπηρεσία Εξετάσεων του Υπουργείου Παιδείας και Πολιτισμού χρησιμοποιεί στατιστικές επεξεργασίες αρχικών βαθμολογιών;

Παρόμοιες (και συνήθως πολύ πιο πολύπλοκες) στατιστικές επεξεργασίες χρησιμοποιούνται από όλες τις ευρέως γνωστές Υπηρεσίες Εξετάσεων του εξωτερικού π.χ. Educational Testing Service (USA), Cambridge Assessment (UK), Cito Group (Netherlands) κτλ.

4. Μόνο στις Εξετάσεις Διορισίμων γίνεται στατιστική επεξεργασία των βαθμολογιών;

Στην Κύπρο εφαρμόζεται η ίδια μέθοδος και για τις Παγκύπριες Εξετάσεις εδώ και δεκαετίες, αλλά και για άλλες εξετάσεις. Η στατιστική επεξεργασία των αρχικών βαθμολογιών δεν είναι απλά μια θεωρία σε ακαδημαϊκά εγχειρίδια, αλλά είναι συμβατή με τις διαδικασίες που ακολουθούνται σε μεγάλες εξετάσεις παγκοσμίως (π.χ. SATs, GRE, ACT).

5. Μπορεί η κατάταξη των υποψηφίων να συνδυαστεί με βαθμολόγηση (π.χ. μοριοδότηση) από άλλες πηγές πληροφόρησης (π.χ. λόγω χρόνων υπηρεσίας);

Κατηγορηματικά ναι. Γιατί όχι; Δέστε για παράδειγμα τις εξετάσεις New South Wales Higher School Certificate (HSC) όπου οι στατιστικά επεξεργασμένες βαθμολογίες από τις γραπτές εξετάσεις συνδυάζονται με βαθμολογίες από άλλες πηγές (π.χ. σχολικούς βαθμούς).

6. Είναι η στατιστική επεξεργασία που χρησιμοποίησε η Υπηρεσία Εξετάσεων πολύπλοκη και αδιαφανής;

Η Υπηρεσία Εξετάσεων χρησιμοποιεί την πιο απλή μέθοδο στατιστικής επεξεργασίας που υπάρχει θεμελιωμένη στη βιβλιογραφία (δέστε σχετικό σχόλιο στο Kolen and Brennan, 2004, σελ. 54).

7. Η μέθοδος στατιστικής επεξεργασίας που χρησιμοποιήθηκε δεν ήταν υπόψη κανενός και υιοθετήθηκε μετά τις εξετάσεις;

Ουδέν αναληθέστερο. Η διαδικασία αυτή ήταν γνωστή, προβλεπόταν από τον σχετικό Νόμο/Κανονισμό και χρησιμοποιείται από την Υπηρεσία Εξετάσεων εδώ και δεκαετίες για τις Παγκύπριες Εξετάσεις μετά από συμβουλές ειδικών από το Educational Testing Service (USA). Η μέθοδος αυτή έχει επίσης παρουσιαστεί εδώ και πολλά χρόνια σε διάφορα επιστημονικά συγγράμματα ως η μέθοδος στατιστικής επεξεργασίας που χρησιμοποιεί η Υπηρεσία Εξετάσεων (δέστε Lamprianou, 2009, 2012). Η Υπηρεσία Εξετάσεων δεν έκανε κάποια νέα επιλογή μεθόδου, απλά χρησιμοποίησε την ίδια μέθοδο (z scores) που χρησιμοποιεί εδώ και δεκαετίες, άρα δεν είχε κάτι καινούριο να ανακοινώσει σχετικά με αυτό το θέμα.

8. *Είναι αλήθεια ότι με τη στατιστική επεξεργασία που χρησιμοποίησε η Υπηρεσία Εξετάσεων αλλάζει η σειρά των επιτυχόντων ανά μάθημα:*

Αναληθέστατο. Δέστε για παράδειγμα το Chang (2006), που αναφέρει ότι η μέθοδος αυτή δεν επηρεάζει το σχήμα της κατανομής των βαθμολογιών («*This transformation process exerts no effect on the original distributional shape of the raw scores*», σελ. 907). Οι ανηγμένες βαθμολογίες είναι απλά η απόσταση της αρχικής βαθμολογίας των υποψηφίων από το μέσο όρο. Δείχνουν την σχετική επιτυχία ενός υποψηφίου σε σχέση με τους ανθυποψηφίους του.

9. *Είναι αλήθεια ότι η στατιστική επεξεργασία σε όλα τα γνωστικά αντικείμενα που εξετάστηκαν οι υποψήφιοι δεν ήταν απαραίτητη;*

Λάθος, δέστε τον Kolen (2006) που συμβουλεύει ότι για να μπορέσουμε να συνθέσουμε τις αρχικές βαθμολογίες των υποψηφίων στα διάφορα αντικείμενα εξέτασης, θα πρέπει να επεξεργαστούμε με την ίδια μέθοδο τις αρχικές βαθμολογίες των υποψηφίων, όπως εξηγήθηκε πιο πριν: «*Using the same scaling procedures for each of the tests in a battery facilitates the formation of ... composites*» (σελ. 169).

10. *Είναι αλήθεια ότι η στατιστική επεξεργασία και η άθροιση των «ανηγμένων» βαθμολογιών πιθανόν – αλλά όχι απαραίτητως – να καθορίζει τη συνολική (όχι ανά μάθημα) κατάταξη των υποψηφίων;*

Ναι, αλλά αυτός ακριβώς είναι ο στόχος της στατιστικής επεξεργασίας επειδή πρέπει να αθροιστούν συγκρίσιμες μετρήσεις από κάθε αντικείμενο εξέτασης προκειμένου να καταταχθούν οι υποψήφιοι. Είναι λάθος να χρησιμοποιηθούν οι ανεπεξέργαστες βαθμολογίες. Δέστε, για παράδειγμα, το σχετικό δημοσίευμα των ερευνητών του Educational Testing Service, Tan και Michel, 2011) ή και τις οδηγίες του Ολλανδικού συστήματος Ποιότητας Εξετάσεων (Dutch Rating System for Test Quality) όπως περιγράφονται στο Evers (2001, σελ. 165) «*Taken by itself, the raw score has little or no*

significance.», δηλαδή από μόνες τους οι αρχικές (ανεπεξέργαστες) βαθμολογίες δεν σημαίνουν τίποτα!

11. Υπάρχουν στη διεθνή ακαδημαϊκή κοινότητα αμφιβολίες ως προς την καταλληλότητα αυτής της μεθόδου;

Κατηγορηματικά όχι. Η Επιστημονική Επιτροπή που επέβλεψε την εφαρμογή της μεθόδου στατιστικής επεξεργασίας για τις Εξετάσεις Διορισίμων έχει μεγάλη εμπειρία και πραγματογνωμοσύνη που έχει κατά καιρούς αξιοποιηθεί από κυβερνήσεις διαφόρων χωρών, Ευρωπαϊκών και μη και από Υπηρεσίες Εξετάσεων σε διάφορες χώρες. Με όλη την εμπειρία από διεθνή συνέδρια και με όλη τη γνώση της σχετικής βιβλιογραφίας, διαπιστώνεται κατηγορηματικά ότι η ακαδημαϊκή κοινότητα δεν έχει αμφιβολίες ως προς τη στατιστική επεξεργασία των αρχικών βαθμολογιών των υποψηφίων. Εάν υπάρχει αντίθετη άποψη, θα πρέπει να δηλωθεί και να τεκμηριωθεί με σχετική βιβλιογραφία σε καταξιωμένα επιστημονικά περιοδικά.

Από τα πιο πάνω προκύπτουν τα ακόλουθα:

1. Ακόμα και αν η στατιστική επεξεργασία δεν προβλεπόταν από το σχετικό Νόμο, η Υπηρεσία Εξετάσεων θα έπρεπε υποχρεωτικά να την υιοθετήσει, προκειμένου η κατάταξη των υποψηφίων να είναι δίκαιη και το κράτος να μπορέσει να καταρτίσει άρτιους καταλόγους διορισίμων, για αξιοποίηση από την Επιτροπή Εκπαιδευτικής Υπηρεσίας.
2. Όσοι έχουν αντίθετη άποψη και πιστεύουν ότι υπάρχει κάποια καλύτερη μέθοδος, θα πρέπει να την παρουσιάσουν και να την τεκμηριώσουν παραθέτοντας τη σχετική επιστημονική βιβλιογραφία.
3. Αν οι επικριτές της μεθόδου στατιστικής επεξεργασίας που χρησιμοποιήθηκε δεν μπορούν να παραθέσουν εξίσου ισχυρά επιχειρήματα, γερά θεμελιωμένα στη διεθνή ακαδημαϊκή γνώση και πρακτική, θα πρέπει άμεσα και δημοσίως να δηλώσουν ότι αποσύρουν τα ανακριβή τους δημοσιεύματα και να απολογηθούν για την αχρείαστη ανησυχία που έχουν προκαλέσει στους ενδιαφερόμενους και στο κοινό. Αν δεν το πράξουν άμεσα, παραμένουν εκτεθειμένοι.

2. Εισαγωγικά

Το θέμα της στατιστικής επεξεργασίας των αποτελεσμάτων ανταγωνιστικών εξετάσεων κατάταξης, έχει απασχολήσει την ακαδημαϊκή κοινότητα παγκοσμίως εδώ και αρκετές δεκαετίες. Το πρόβλημα έγκειται στο γεγονός ότι οι αρχικές βαθμολογίες, δηλαδή οι μη επεξεργασμένοι βαθμοί που αναγράφονται πάνω στα γραπτά των υποψηφίων, δεν προσφέρονται για να

καταδείξουν την επιτυχία ενός υποψηφίου σε σχέση με τους ανθυποψηφίους, όταν αυτοί διαγωνίζονται σε περισσότερα του ενός αντικείμενα. Για παράδειγμα, αν σε μίαν ανταγωνιστική εξέταση κατάταξης κάποιος πέτυχε αρχική βαθμολογία 80%, ο βαθμός του αυτός πρέπει να συγκριθεί με τις βαθμολογίες των υπόλοιπων υποψηφίων, προκειμένου να σχηματίσουμε σαφή εικόνα της επιτυχίας ή της αποτυχίας του (πιθανόν οι περισσότεροι υποψήφιοι να εξασφάλισαν βαθμολογίες πέραν του 80%). Σε ένα άλλο γνωστικό αντικείμενο της ίδιας εξέτασης, ένας υποψήφιος μπορεί να εξασφάλισε βαθμολογία 60% και να θεωρηθεί μεγάλη επιτυχία, επειδή οι περισσότεροι υποψήφιοι εξασφάλισαν χαμηλότερες βαθμολογίες. Πώς, λοιπόν, συνθέτουμε τα αποτελέσματα των εξετάσεων, όταν οι υποψήφιοι εξετάζονται σε γνωστικά αντικείμενα διαφορετικής δυσκολίας; Ευτυχώς οι ερευνητές ανέπτυξαν συγκεκριμένες στατιστικές τεχνικές για να αντιμετωπίσουν αυτό το πρόβλημα.

Προκειμένου να συσσωρευτεί η σύγχρονη συλλογική σοφία της ψυχομετρικής επιστημονικής κοινότητας σε ένα εγχειρίδιο, το Εθνικό Συμβούλιο Εκπαιδευτικής Μέτρησης των Ηνωμένων Πολιτειών (National Council on Measurement in Education) και το Αμερικανικό Συμβούλιο Εκπαίδευσης (American Council on Education), χρηματοδότησαν ένα συλλογικό τόμο σχεδόν 800 σελίδων (Brennan, 2006) ο οποίος και αποτελεί έκτοτε σημείο αναφοράς για τους ερευνητές και όσους παίρνουν αποφάσεις για θέματα εξετάσεων κατάταξης. Η Υπηρεσία Εξετάσεων του Υπουργείου Παιδείας αντλεί και εφαρμόζει μεθόδους στατιστικής επεξεργασίας βαθμολογιών που περιγράφονται σε αυτό τον τόμο. Βασική αρχή είναι ότι οι αρχικές βαθμολογίες από εξετάσεις διαφορετικών αντικειμένων δεν πρέπει να αθροίζονται όταν θα συγκριθούν οι υποψήφιοι μεταξύ τους. Χαρακτηριστικά, ο Kolen (2006, σελ. 155) λέει ότι οι εξεταστές (test developers) προσδίδουν νόημα στις αρχικές βαθμολογίες μέσω των «ανηγμένων βαθμολογιών» (scaled scores), προκειμένου οι χρήστες (test users) να μπορέσουν να αξιοποιήσουν τα αποτελέσματα για σκοπούς κατάταξης των υποψηφίων.

Σχετικά με τη στατιστική επεξεργασία αρχικών βαθμολογιών υπάρχει ομοφωνία των ερευνητών: Όπως αναφέρουν οι Tan και Michel (2011), εκφράζοντας τη θέση του Educational Testing Service (USA), πολύ απλά οι αρχικές (μη επεξεργασμένες) βαθμολογίες δεν μπορούν να χρησιμοποιηθούν για να συγκριθεί η επίδοση υποψηφίων σε διαφορετικά τεστ (*«the raw scores cannot be used to compare test takers' performances on different forms»*, σελ. 2). Το ίδιο προβλέπεται και στις οδηγίες του Ευρωπαϊκού Συνδέσμου Εκπαιδευτικής Αξιολόγησης (Association for Educational Assessment – Europe, δέστε Standards for Assessment, σελ 11) *«In the case of a series of assessments, measures should be taken to store the outcomes on the various assessments in such a way as to allow the aggregation of results»*. Το ίδιο προβλέπεται

επίσης και στις οδηγίες της Διεθνούς Επιτροπής Εξετάσεων (International Test Commission, παράγραφος 2.6.2, σελ. 20), «*Carry out appropriate raw score conversions to other relevant types of scale*». Τέλος, οι οδηγίες του Ολλανδικού συστήματος Ποιότητας Εξετάσεων (Dutch Rating System for Test Quality), όπως περιγράφονται στο Evers (2001, σελ. 165) είναι σαφέστατες: «*Taken by itself, the raw score has little or no significance*», δηλαδή από μόνες τους οι αρχικές (ανεπεξεργαστες) βαθμολογίες δεν σημαίνουν τίποτα!

3. Οι Εξετάσεις Διορισίμων και η Διεθνής Βιβλιογραφία

Στις πρόσφατες Εξετάσεις Διορισίμων, το Υπουργείο Παιδείας και Πολιτισμού και η Υπηρεσία Εξετάσεων είχαν δύο σημαντικές προκλήσεις να αντιμετωπίσουν: την κατάταξη των υποψηφίων για την τρέχουσα εξεταστική περίοδο 2017-2018 και τη μοριοδότηση των επιτυχόντων υποψηφίων με τρόπο που να επιτρέπει τη συγκρισιμότητα των μορίων με τους υποψηφίους που θα επιλέξουν να παρακαθίσουν στις εξετάσεις της ίδιας ειδικότητας κατά τις επόμενες εξεταστικές περιόδους. Να σημειωθεί εδώ ότι οι δύο προκλήσεις είναι στενά αλληλένδετες και δεν είναι δυνατό να αντιμετωπιστεί η μια αδιαφορώντας για την άλλη.

Για την αντιμετώπιση του προβλήματος της κατάταξης των υποψηφίων η Υπηρεσία Εξετάσεων επέλεξε να χρησιμοποιήσει την πιο απλή μέθοδο στατιστικής επεξεργασίας που υπάρχει (δείτε σχετικό σχόλιο στο Kolen and Brennan, 2004, σελ. 54, αλλά και εκτεταμένες αναφορές στο Brennan, 2006). Η μέθοδος αυτή περιγράφεται στην ανακοίνωση της Υπηρεσίας Εξετάσεων σχετικά με τη Στατιστική Επεξεργασία των βαθμολογιών των Εξετάσεων Διορισίμων και είναι διαθέσιμη στο κοινό στο σύνδεσμο http://archeia.moec.gov.cy/ed/117/2018_03_30_statistiki_epexergasia_vathmologion.pdf. Η μέθοδος αυτή περιγράφεται επίσης – πιο συμπυκνωμένα - και στον Οδηγό Εξετάσεων των Παγκύπριων Εξετάσεων (Τόμος I), επειδή το Υπουργείο Παιδείας και Πολιτισμού υιοθέτησε εδώ και δεκαετίες την ίδια μέθοδο στατιστικής επεξεργασίας και για τις Παγκύπριες Εξετάσεις, μετά από εισήγηση του Educational Testing Service (USA). Αν και πολύ απλή, η μέθοδος αυτή έχει κριθεί ως εξαιρετικά χρήσιμη από την ακαδημαϊκή κοινότητα (π.χ. δες την αξιολόγηση του Chang, 2006) και είναι επίσης γνωστή και ως linear scaling ή z-scores.

Σύμφωνα με τη σχετική βιβλιογραφία, ο εξεταστής (στην περίπτωσή μας η Υπηρεσία Εξετάσεων), καθορίζει τον μέσο όρο και την τυπική απόκλιση των «*ανηγμένων βαθμολογιών*» (scaled scores) αναλόγως των αναγκών του και των στόχων της εξέτασης. Για τη στατιστική επεξεργασία, χρησιμοποιούνται, επίσης, ο μέσος όρος και η τυπική απόκλιση των αρχικών βαθμολογιών της ομάδας αναφοράς (reference group) που στην περίπτωση των Εξετάσεων Διορισίμων είναι η

ομάδα των επιτυχόντων σε κάθε μάθημα (όχι όλοι οι υποψήφιοι). Η χρήση του όρου «ομάδα αναφοράς» είναι πολύ σημαντική, επειδή υπονοεί ότι όλες οι επόμενες εξετάσεις θα «εξισώνουν» τις μοριοδοτήσεις των επόμενων εξεταστικών περιόδων με αυτές της ομάδας αναφοράς, προκειμένου όλοι να βρίσκονται στον ίδιο κατάλογο και να συγκρίνονται δίκαια. Αυτή είναι και η δηλωμένη πρόθεση του Νομοθέτη, εφόσον προβλέπεται ότι οι επιτυχόντες «διατηρούν» τα μόριά τους, εφόσον το επιθυμούν, μέχρι δέκα χρόνια.

Σχετικά με την πρακτική εφαρμογή αυτής της μεθόδου, στον Kolen (2006, σελ. 163), παρουσιάζεται ένα επεξεργασμένο παράδειγμα: έστω μια υποθετική εξέταση, όπου ο μέσος όρος των αρχικών βαθμολογιών είναι 70 και η τυπική απόκλιση είναι 10 και οι εξεταστές επιθυμούν να μετασχηματίσουν τις βαθμολογίες σε μια «ανηγμένη κλίμακα» με μέσο όρο 20 και τυπική απόκλιση 5. Σε αυτή την περίπτωση, ο επεξεργασμένος βαθμός ενός υποψηφίου με αρχική βαθμολογία 50 θα είναι 10, ενώ ένας υποψήφιος με αρχική βαθμολογία 60 θα έχει επεξεργασμένο βαθμό 15. Στα μάτια των υποψηφίων, η αρχική διαφορά των $70-60=10$ ανεπεξέργαστων μονάδων έχει γίνει $15-10=5$ μονάδες «ανηγμένης» βαθμολογίας και ενδεχομένως κάποιος να διαμαρτυρηθεί. Στην πραγματικότητα δεν υπάρχει λόγος διαμαρτυρίας, επειδή δεν έχει συμβεί καμιά αδικία: οι επεξεργασμένες βαθμολογίες απλά απεικονίζουν την πραγματική απόσταση της κατάταξης (*relative standing*) των υποψηφίων στη νέα κλίμακα μέτρησης. Με τη μέθοδο αυτή, ο μετασχηματισμός των βαθμολογιών ανά εξέταση είναι γραμμικός και η σειρά των υποψηφίων δεν αλλάζει. Επίσης, οι αποστάσεις μεταξύ των βαθμολογιών των υποψηφίων στην κάθε εξέταση (στο κάθε αντικείμενο) διατηρούνται αναλογικά, όπως ακριβώς ισχυρίζεται και στην ανακοίνωσή της η Υπηρεσία Εξετάσεων. Όπως λέει και ο Chang (2006), η διαδικασία αυτή δεν επηρεάζει το σχήμα της κατανομής των βαθμολογιών (*«This transformation process exerts no effect on the original distributional shape of the raw scores»*, σελ. 907). Και στο σημείο αυτό, λοιπόν, η Υπηρεσία Εξετάσεων έχει απλά υιοθετήσει θεμελιωμένες τεχνικές, όπως αυτές περιγράφονται στη σχετική βιβλιογραφία.

Στην περίπτωση των Εξετάσεων Διορισίμων, οι υποψήφιοι εξετάζονται σε τρία αντικείμενα :

- εξέταση στο αναλυτικό πρόγραμμα του γνωστικού αντικειμένου (32%)
- εξέταση δεξιοτήτων που πρέπει να χαρακτηρίζουν τους εκπαιδευτικούς (10%)
- εξέταση γνώσης της Ελληνικής γλώσσας (8%)

Συνεπώς, μιλούμε για ομάδα εξεταστικών δοκιμίων (*battery of tests*). Ο Kolen (2006) αναφέρει ότι για να μπορέσουμε να συνθέσουμε τις αρχικές βαθμολογίες των υποψηφίων από τα διάφορα αντικείμενα, θα πρέπει να επεξεργαστούμε με την ίδια μέθοδο τις αρχικές βαθμολογίες των υποψηφίων, όπως εξηγήθηκε πιο πάνω: *«Using the same scaling procedures for each of the*

tests in a battery facilitates the formation of ... composites» (σελ. 169). Όπως αναφέρουν οι Tan και Michel (2011), εκφράζοντας και τη θέση του Educational Testing Service, πολύ απλά οι αρχικές (μη επεξεργασμένες) βαθμολογίες δεν μπορούν να χρησιμοποιηθούν για να συγκριθεί η επίδοση υποψηφίων σε διαφορετικές εξετάσεις («the raw scores cannot be used to compare test takers' performances on different forms», σελ. 2).

Αυτό ακριβώς κάνει και η Υπηρεσία Εξετάσεων: Το αποτέλεσμα αυτής της διαδικασίας είναι ότι η «σειρά» των υποψηφίων είναι δυνατόν, αλλά όχι απαραίτητο, να αλλάξει (σε σχέση με το άθροισμα των αρχικών τους βαθμολογιών). Αυτό όμως δεν είναι παρενέργεια, λάθος ή αβλεψία αλλά η ουσία της διαδικασίας: οι αρχικές βαθμολογίες σε κάθε εξέταση πρέπει να μετασχηματίζονται σε «ανηγμένες» και αυτές να συνθέτονται (π.χ. να προστίθενται), ώστε να ποσοτικοποιηθεί η συνολική ικανότητα του κάθε υποψηφίου («... *making possible the assessment of examinee strength and weaknesses across test areas»*, σελ. 169). Είναι κατανοητό κάποιιο υποψήφιο να παίρνουν την υπολογιστική τους μηχανή και κάνοντας απλές προσθέσεις να ισχυρίζονται ότι «άλλαξε η σειρά επιτυχίας των υποψηφίων εξαιτίας της στατιστικής επεξεργασίας». Είναι κατανοητό να διαμαρτύρονται καθώς δε γνωρίζουν τα ειδικά χαρακτηριστικά μιας τέτοιας διαδικασίας. Αυτό, ωστόσο, που πρέπει να αποδεχτούν είναι ότι η διαδικασία αυτή στην περίπτωση των Εξετάσεων Διορισίμων ήταν μονόδρομος και η Υπηρεσία Εξετάσεων ακολούθησε ευλαβικά μία ευρύτατα αποδεκτή από τη διεθνή επιστημονική κοινότητα, διαδικασία.

4. Η Υπηρεσία Εξετάσεων και οι Διεθνείς Πρακτικές

Οι διαδικασίες που χρησιμοποιεί η Υπηρεσία Εξετάσεων δεν είναι απλά θεωρίες σε ακαδημαϊκά εγχειρίδια, αλλά είναι συμβατές με τις διαδικασίες που ακολουθούνται σε μεγάλες εξετάσεις παγκοσμίως (π.χ. SATs, GRE, ACT). Για παράδειγμα, οι εξετάσεις ACT στις ΗΠΑ, αντί να χρησιμοποιήσουν τις αρχικές βαθμολογίες των υποψηφίων, χρησιμοποιούν ανηγμένες (η μέθοδος στατιστικής επεξεργασίας που χρησιμοποιούν είναι πολύπλοκη). Η μέγιστη βαθμολογία που μπορεί κάποιος να εξασφαλίσει στις εξετάσεις ACT είναι το 36 (στις εξετάσεις Διορισίμων είναι το 50). Στις εξετάσεις ACT αυτός ο βαθμός προκύπτει από τον μέσο όρο των ανηγμένων βαθμολογιών κάθε υποψηφίου σε τέσσερα τεστ (στις εξετάσεις Διορισίμων προκύπτει από τον μέσο όρο των ανηγμένων βαθμολογιών σε τρία τεστ). Ενώ στις εξετάσεις ACT δεν ανακοινώνεται στους υποψηφίους η αρχική τους βαθμολογία, στις εξετάσεις Διορισίμων όλες οι αρχικές βαθμολογίες έχουν ανακοινωθεί από την Υπηρεσία Εξετάσεων σε κάθε υποψήφιο για σκοπούς διαφάνειας.

Ο εξεταστής (δηλαδή η Υπηρεσία Εξετάσεων), έχει τη δυνατότητα να ορίσει κατά τον ακριβή μετασχηματισμό για κάθε εξέταση (δηλαδή τον μέσο όρο και την τυπική απόκλιση) και να προσδώσει τις κατάλληλες βαρύτητες που να εξυπηρετούν τους σκοπούς της εξέτασης. Για παράδειγμα, στην περίπτωση των αμερικάνικων εξετάσεων SATs, οι κλίμακες της Λεκτικής και Μαθηματικής ικανότητας έχουν μέσο όρο 500 και τυπική απόκλιση 110 (Dorans, 2002). Ωστόσο, σε μια άλλη περίπτωση, για παράδειγμα, για το ACT Assessment, κάθε «ανηγμένη κλίμακα» έχει μέσο όρο 18 (Brennan, 1989).

Προκειμένου να εξυπηρετήσει τους σκοπούς των Εξετάσεων Διορισίμων, ο Νομοθέτης προνόησε το κάθε αντικείμενο εξέτασης να έχει διαφορετική βαρύτητα, άρα διαφορετικό μέσο όρο και διαφορετική τυπική απόκλιση αφού επιθυμούμε να εκφράσουμε την επίδοση των υποψηφίων ανά γνωστικό αντικείμενο σε διαφορετική κλίμακα μορίων (π.χ. για την εξέταση Ειδικότητας προβλέπονται 32 μόρια ενώ για τη Γνώση Ελληνικής Γλώσσας προβλέπονται 8). Αυτό είναι στα πλαίσια των δικαιωμάτων του εξεταστή (της Υπηρεσίας Εξετάσεων) που έχει επιβαρυνθεί με το καθήκον να διεκπεραιώσει τις Εξετάσεις Διορισίμων.

Κάποιες φορές οι ανηγμένες βαθμολογίες υποψηφίων από γραπτές εξετάσεις, συνδυάζονται με άλλες βαθμολογίες από άλλες πηγές πληροφόρησης. Για παράδειγμα, στην Αυστραλία, οι εξετάσεις New South Wales Higher School Certificate (HSC) συνδυάζουν ανηγμένες βαθμολογίες από γραπτές εξετάσεις, με βαθμολογίες από άλλη αξιολόγηση, ώστε να υπολογιστεί για κάθε υποψήφιο ένας συνολικός «ανηγμένος» βαθμός. Η ακριβής διαδικασία είναι αρκετά πολύπλοκη αλλά η αναλογία με τις Εξετάσεις Διορισίμων είναι φανερή: στις Εξετάσεις Διορισίμων οι ανηγμένες βαθμολογίες (δηλαδή τα μόρια από τις εξετάσεις), συνδυάζονται με μοριοδότηση από άλλες πηγές (π.χ. προσόντα, χρόνια υπηρεσίας κτλ). Ούτε σε αυτό τον τομέα η Υπηρεσία Εξετάσεων έχει πρωτοτυπήσει με κάποιο παράδοξο τρόπο (εξάλλου η πρόνοια περί μοριοδότησης 50% από τις εξετάσεις και 50% από άλλες πηγές επιβάλλεται από τον Νόμο και πρόκειται για ανεξάρτητες διαδικασίες).

5. Οι αντιδράσεις

Όλα όσα έχουν αναφερθεί πιο πάνω δεν είναι καινούρια. Εφαρμόζονται εδώ και δεκαετίες στην περίπτωση των Παγκύπριων Εξετάσεων, αλλά επίσης εφαρμόζονται και από Υπηρεσίες Εξετάσεων πολλών άλλων χωρών και ασφαλώς δεν είναι ανακάλυψη του Υπουργείου Παιδείας και Πολιτισμού της Κύπρου. Ο αναγνώστης που ενδιαφέρεται να μάθει περισσότερα για συγκεκριμένα παραδείγματα χωρών που χρησιμοποιούν παρόμοια ή ακόμα και πολύ πιο πολύπλοκα συστήματα στατιστικής επεξεργασίας από αυτό που χρησιμοποιούμε στην Κύπρο

μπορεί να ανατρέξει στο Lamprianou (2009). Τίποτα από όσα ακούγονται γύρω μας αυτές τις μέρες στην Κύπρο δεν είναι καινούριο ή απρόσμενο. Η κοινωνία και οι υποψήφιοι έχουν το δικαίωμα να εκφράζουν κόσμια τις απορίες τους και είναι καθήκον του Υπουργείου Παιδείας και Πολιτισμού να επεξηγήσει επιστημονικά το θέμα, τονίζοντας ότι οι διαδικασίες που εφαρμόστηκαν έγιναν αξιοκρατικά έχοντας ως γνώμονα το καλώς νοούμενο συμφέρον της εκπαίδευσης και κατ'επέκταση της κυπριακής κοινωνίας. Θα πρέπει να υπάρξει ενημέρωση της κοινωνίας ώστε να αποφεύγονται άσκοπες αντιδράσεις οι οποίες οφείλονται, θέλουμε να πιστεύουμε, σε άγνοια των σχετικών διαδικασιών. Όπως πολύ εύστοχα είπε, με γνήσιο αγγλοσαξωνικό χιούμορ, εκπρόσωπος της Επιτροπής για τη Στατιστική Επεξεργασία βαθμολογιών στην Αυστραλία (Technical Committee on Scaling, 2002) *'each year brings its own myths and conspiracy theories'* (κάθε χρονιά έχουμε καινούριους μύθους και θεωρίες συνωμοσίας)!

Ο κύπριος πολίτης δεν πρέπει να έχει αμφιβολίες για τη στατιστική επεξεργασία που έχει χρησιμοποιηθεί για τις Εξετάσεις Διορισίμων: Όλες οι διαδικασίες που έχουν εφαρμοστεί είναι καλά θεμελιωμένες στη σχετική διεθνή βιβλιογραφία και πρακτική.

6. Συγκρισιμότητα με μελλοντικές εξετάσεις

Η δεύτερη πρόκληση που είχε να αντιμετωπίσει η Υπηρεσία Εξετάσεων πηγάζει από την πρόνοια του Νομοθέτη ο υποψήφιος να «διατηρεί», αν το επιθυμεί, τα μόρια που εξασφάλισε μέχρι και δέκα χρόνια. Η μέθοδος στατιστικής επεξεργασίας που εφαρμόστηκε, επιτρέπει την κατάταξη των υποψηφίων για την τρέχουσα εξεταστική περίοδο, δεν διασφαλίζει όμως ότι η μοριοδότησή τους θα είναι συγκρίσιμη με τη μοριοδότηση των επόμενων εξεταστικών περιόδων.

Η Υπηρεσία Εξετάσεων έχει προνοήσει γι' αυτό και όπως αναφέρει στην ανακοίνωσή της, το «Στάδιο 5» της διαδικασίας που θα εφαρμοστεί στις επόμενες εξετάσεις θα καλύψει τις αδυναμίες που παραμένουν. Οι λεπτομέρειες της μεθόδου αυτής, (το «Στάδιο 5»), δεν έχουν προφανώς ακόμα αποφασιστεί, επειδή έπρεπε πρώτα να γίνουν διαθέσιμα τα αποτελέσματα της πρώτης χρονιάς, προκειμένου να αξιολογηθούν και να επιλεγεί η καλύτερη τεχνική λύση.

Για τον σκοπό αυτό, το Υπουργείο Παιδείας και Πολιτισμού έχει ήδη ενδυναμώσει την υφιστάμενη ομάδα εμπειρογνομόνων με ακαδημαϊκούς από το εξωτερικό και υπάρχει αρκετός χρόνος για τις σχετικές μελέτες (περίπου 18 μήνες) και θα θέσει στη διάθεση της Επιστημονικής Επιτροπής όλους τους αναγκαίους πόρους.

Θα ήταν ευχής έργο, επίσης, αν μπορούσαμε να είχαμε περισσότερα δεδομένα στη διάθεσή μας, (π.χ. αποτελέσματα από 5-6 εξεταστικές περιόδους), προκειμένου να αξιολογήσουμε όλες τις

δυνατότητές μας και να κάνουμε τις πιο σωστές επιλογές. Βρισκόμαστε όμως στα πρώτα βήματα, οι Εξετάσεις Διορισίμων έχουν «τρέξει» μόνο μια φορά. Θα επιδιώξουμε τις βέλτιστες επιλογές αξιολογώντας τα δεδομένα που διαθέτουμε με βάση τη διεθνή βιβλιογραφία και τη διεθνή πρακτική.

7. Τα επόμενα βήματα

Η κυπριακή κοινωνία, μέσω των εκπροσώπων της στη Βουλή, αποφάσισε να πραγματοποιήσει τις Εξετάσεις Διορισίμων, προκειμένου να επιλέγονται οι άριστοι. Αυτό θα ωφελήσει το εκπαιδευτικό σύστημα και τα παιδιά μας, ωστόσο η υλοποίηση της σχετικής Νομοθεσίας είναι φυσικό να απαιτεί πόρους, χρόνο και τον σχεδιασμό και την υλοποίηση νέων διαδικασιών.

Οι ενδιαφερόμενοι, (π.χ. υποψήφιοι, ομάδες πίεσης), πρέπει να γνωρίζουν ότι δεν υπάρχουν τέλειες τεχνικές λύσεις σε πολύπλοκα προβλήματα, ειδικότερα όταν αυτά αντιμετωπίζονται για πρώτη φορά.

8. Συμπερασματικά

Η Υπηρεσία Εξετάσεων έχει υιοθετήσει θεμελιωμένες τεχνικές στατιστικής επεξεργασίας, όπως αυτές προβλέπονται στη σχετική επιστημονική βιβλιογραφία και τη διεθνή πρακτική.

Συνεπώς, είναι λάθος να διαβρώνεται ασκόπως και αδίκως η εμπιστοσύνη του κοινού για τις διαδικασίες που εφαρμόστηκαν από την Υπηρεσία Εξετάσεων, έναν θεσμό που υπηρετεί τον Κύπριο πολίτη για δεκαετίες με αξιοπρέπεια και εντιμότητα.

Κατάλογος αναφορών

Brennan, R. L. (Ed.) (1989). *Methodology used in scaling the ACT Assessment and P-ACT+*. Iowa City, Iowa: ACT Publications.

Brennan, L. R. (Ed.) (2006). *Educational Measurement*. (4th edition) New York: American Council on Education/Praeger.

Dorans, N. J. (2002). Recentering and realigning the SAT score distributions: How and why? *Journal of Educational Measurement*, 39(1), 59-84.

Evers, A. (2001). The Revised Dutch Rating System for Test Quality, *International Journal of Testing*, 1:2, 155-182.

International Test Commission (2001). International Guidelines for Test Use, *International Journal of Testing*, 1(2), 93-114.

Kolen, M. J. (2006). Scaling and Norming. In Brennan, R. L. (ed.), *Educational Measurement*. (4th edition) New York: American Council on Education/Praeger, pp. 155-186.

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling and Linking*. (2nd edition). Springer.

Lamprianou, I. (2009). Comparability of Examination Standards Between Subjects: an International Perspective. *Oxford Review of Education*, 35 (2), 205-226.

Lamprianou, I. (2012). Effects of Forced Policy-Making in High Stakes Examinations: the Case of the Republic of Cyprus. *Assessment in Education: Principles, Policy & Practice*, 19 (1), 27-44.

Newton, P. E. (2013). Ofqual's Reliability Programme: a case study exploring the potential to improve public understanding and confidence. *Oxford Review of Education*, 39(1), 93-113.

Tan, X., & Michel, R. (2011). *Why Do Standardized Testing Programs Report Scaled Scores?* R & D Connections, Educational Testing Service.

Technical Committee on Scaling (2002). *Report of calculation of universities admission index 2001, for Committee of Chairs*. Available online at: http://www.secretariat.unsw.edu.au/acboard/committee_chairs/tcsrep302a.pdf (accessed 12 March 2007).

The Association of Educational Assessment – Europe (2012). *European Framework of Standards for Educational Assessment 1.0*. AEA-E, Roma: Edizioni Nuova Cultura.