

Αξιολόγηση της εγκυρότητας κριτηρίου του Κρατικού Πιστοποιητικού Γλωσσομάθειας της Αγγλικής Γλώσσας

Μαρία Τσομπάνογλου
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
mariatsompanoglou@yahoo.co.uk.

Abstract: The present article aims at the criterion-related validation of the Greek State Certificate of English Language Proficiency (B2 level)¹. Two criterion-tests were used: QPT² and FCE³.

The scores of 124 subjects on KPG and the two test-criteria were analyzed through SPSS 11.0 using Pearson correlation coefficient. The evidence collected led to the confirmation of the criterion-related validity of KPG.

The subjects' total scores on KPG showed positive correlations with their scores on both test-criteria. The validity of KPG was further confirmed by the positive correlations of subjects' sub-scores on the different exam papers of KPG and FCE.

Key words: KPG, test validation, criterion-related validity, concurrent validity, external validity, empirical validity, statistical validity.

1. Εισαγωγή

Η εγκυρότητα αποτελεί το θεμέλιο λίθο της γλωσσικής, αλλά και κάθε είδους αξιολόγησης. Γι' αυτό το λόγο κάθε 'σοβαρό' μέσο γλωσσικής αξιολόγησης οφείλει να παρουσιάσει μια ικανοποιητική απόδειξη της εγκυρότητάς του από όσο το δυνατόν περισσότερες πηγές, αν θέλει τα αποτελέσματά του να έχουν ισχύ.

Ωστόσο, παρά το γεγονός ότι οι περισσότεροι 'κατασκευαστές' μέσων γλωσσικής αξιολόγησης αναγνωρίζουν τη σπουδαιότητα της διασφάλισης της εγκυρότητάς τους, ελάχιστοι είναι αυτοί που όντως το αποδεικνύουν διεξάγοντας έρευνες πάνω στην εγκυρότητα των 'δημιουργημάτων' τους. Όσον αφορά τα συστήματα πιστοποίησης γλωσσομάθειας, η αλήθεια είναι ότι πολύ λίγα διεξάγουν και δημοσιεύουν έρευνες για την εγκυρότητά τους.

Στόχος του άρθρου είναι η αξιολόγηση της εγκυρότητας του Κρατικού Πιστοποιητικού Γλωσσομάθειας του επιπέδου B2 για την αγγλική γλώσσα⁴. Το γεγονός ότι αποτελεί ένα σχετικά καινούργιο και ευρέως χρησιμοποιούμενο τεστ επάρκειας της αγγλικής γλώσσας 'επιβάλλει' την ανάγκη ελέγχου της εγκυρότητάς του.

Το βασικό ερώτημα που θα πρέπει να τεθεί στον έλεγχο της εγκυρότητας ενός τεστ⁵ είναι αν το τεστ όντως εξετάζει αυτό που υποστηρίζει ότι εξετάζει. Η απάντηση, ωστόσο, στο παραπάνω ερώτημα δεν είναι ούτε απλή, ούτε μονόπλευρη. Αντίθετα, απαιτεί τη συλλογή μιας σειράς αποδείξεων, καθεμία από τις οποίες στοχεύει και σε μια διαφορετική της πλευρά.

¹ The test is part of a Greek multi-level standardized suite of national language examinations, known as 'Kratiko Pistopitiko Glossomathias' or KPG.

² Quick Placement Test.

³ First Certificate in English.

⁴ Το Κρατικού Πιστοποιητικό Γλωσσομάθειας του επιπέδου B2 για την αγγλική γλώσσα θα αναφέρεται στο εξής ως ΚΠΓ.

⁵ Ο όρος *τεστ* χρησιμοποιείται στο παρόν άρθρο με την έννοια του μέσου γλωσσικής αξιολόγησης γενικότερα.

Η παρούσα μελέτη επικεντρώνεται σε μία ιδιαίτερα σημαντική πλευρά της εγκυρότητας του ΚΠΓ, την εγκυρότητα κριτηρίου του και συγκεκριμένα τη σύγχρονη εγκυρότητά του. Σύμφωνα με το *Multilingual Glossary of Language Testing Terms* (ALTE, 1998), ως εγκυρότητα ενός τεστ ορίζεται «ο βαθμός στον οποίο τα αποτελέσματα ενός τεστ μας οδηγούν σε κατάλληλα, με νόημα, και χρήσιμα συμπεράσματα δεδομένου του σκοπού του τεστ». Η έμφαση που δίνεται στα *αποτελέσματα* ενός τεστ είναι προφανής και αυτά ακριβώς αποτελούν το αντικείμενο ανάλυσης στον έλεγχο της εγκυρότητας κριτηρίου του, όπως εξηγείται και παρακάτω.

2. Εγκυρότητα κριτηρίου

Η *εγκυρότητα κριτηρίου* (criterion-related validity), ή αλλιώς, σύμφωνα με τον Heaton (1975: 154), *εμπειρική* (empirical validity) ή *στατιστική εγκυρότητα* (statistical validity) βασίζεται στη σύγκριση του τεστ με μια *εξωτερική μεταβλητή* (external variable), η οποία αποτελεί το *κριτήριο* (criterion). Γι αυτό και σύμφωνα με τους (Palmer et al. 1981: 2), λέγεται και *εξωτερική εγκυρότητα* (external validity). Κατά τον Messick (1995), αντιστοιχεί στην *εξωτερική πλευρά* (external aspect) της συνολικής εγκυρότητας ενός τεστ, καθώς βασίζεται στη συσχέτισή του με εξωτερικά κριτήρια.

Αυτού του είδους η εγκυρότητα μπορεί να ελεγχθεί συγκρίνοντας τα αποτελέσματα του υπό αξιολόγηση τεστ με τα αποτελέσματα κάποιου άλλου μέτρου που θα λειτουργήσει ως κριτήριο, είτε ταυτόχρονα, είτε υστερόχρονα. Ανάλογα με το αν το κριτήριο είναι ταυτόχρονο με το τεστ ή υστερόχρονο, η εγκυρότητα κριτηρίου είναι *σύγχρονη* ή *συνυπάρχουσα* (concurrent validity) ή *προγνωστική* ή *προβλεπτική* (predictive validity), αντίστοιχα.

Στη σύγχρονη εγκυρότητα, η οποία και θα εξεταστεί στην παρούσα έρευνα, το κριτήριο μπορεί να αποτελέσει ένα ήδη υπάρχον τεστ το οποίο να θεωρείται έγκυρο, ή ακόμη και η βαθμολογία του/των καθηγητών ή κάποια άλλη μορφή ανεξάρτητης αξιολόγησης, όπως για παράδειγμα η *αυτοαξιολόγηση* (self-assessment). Σημαντική θετική συσχέτιση μεταξύ των αποτελεσμάτων στα δύο τεστ αποτελεί ένδειξη εγκυρότητας.

Αντίθετα, στην προβλεπτική εγκυρότητα το κριτήριο μπορεί να είναι είτε η μετέπειτα απόδοση των υποψήφιων σε κάποιο άλλο έγκυρο τεστ, είτε η μετέπειτα αξιολόγησή τους από τους καθηγητές τους, ή κάποια άλλη μορφή ανεξάρτητης αξιολόγησης. Μέσω αυτού του είδους εγκυρότητας εξετάζεται η 'δύναμη' ενός τεστ να 'προβλέψει' τη μελλοντική επιτυχία ή αποτυχία των υποψήφιων. Εφόσον, δηλαδή, τα αποτελέσματα στο τεστ συμφωνούν με το μελλοντικό κριτήριο σύγκρισης, το τεστ διαθέτει προβλεπτική εγκυρότητα. Ωστόσο, αυτού του είδους η εγκυρότητα δεν προτιμάται τόσο λόγω πρακτικών προβλημάτων όσο και λόγω της πιθανής μεσολάβησης παραγόντων άσχετων με την ή τις ικανότητες που αξιολογούνται.

Στη σύγχρονη εγκυρότητα αυτό που έχει σημασία είναι η καταλληλότητα του εξωτερικού κριτηρίου. Καταρχήν το τεστ-κριτήριο θα πρέπει να είναι ένα έγκυρο μέτρο της ιδιότητας που εξετάζεται από το τεστ υπό αξιολόγηση, όπως αναφέρουν οι Bachman (1990: 248) και Weir (2005: 36). Γενικότερα, θα πρέπει να είναι ένα 'καλό' τεστ. Σύμφωνα με τους Hatch και Farhady (1982), ένα 'καλό' τεστ πρέπει να έχει τρία χαρακτηριστικά: αξιοπιστία, εγκυρότητα, αλλά και πρακτικότητα.

Σύμφωνα με τον Weir (1988, 2005), το τεστ-κριτήριο συνήθως είναι ένα 'αρχαιότερο' και αναγνωρισμένο τεστ, το οποίο έχει καθιερωθεί. Αλλά και για τον Brown (2000: 1), το τεστ-κριτήριο θα πρέπει να είναι ένα μέτρο που να έχει κερδίσει την εκτίμηση αυτών που το χρησιμοποιούν.

Ωστόσο, θα πρέπει να είναι και πρακτικό. Γι' αυτό το λόγο, κατά την επιλογή ενός τεστ-κριτηρίου, λογικό θα ήταν να λαμβάνονται υπόψη όλα τα πιθανά πρακτικά προβλήματα που θα μπορούσαν να προκύψουν.

3. Κρατικό Πιστοποιητικό Γλωσσομάθειας

Το Κρατικό Πιστοποιητικό Γλωσσομάθειας ανήκει στο *Διεθνές Δίκτυο Συστημάτων Πιστοποίησης* (International Certificate Conference ή ICC) και η δημιουργία του βασίζεται στους *περιγραφικούς δείκτες* (descriptors) του *Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς* (Common European Framework of Reference ή CEFR).

Η εξέταση του επιπέδου B2 συνίσταται από δύο στάδια: τη Γραπτή και την Προφορική Εξέταση. Το πρώτο στάδιο αποτελείται από τρεις φάσεις ή αλλιώς *εξεταστικές ενότητες* (papers).

Στην πρώτη ενότητα (Paper 1) ελέγχεται η *Κατανόηση Γραπτού Λόγου* (Reading Comprehension), καθώς και η *Γλωσσική Συνειδητότητα* (Language Awareness)⁶ του υποψήφιου. Η δεύτερη εξεταστική ενότητα (Paper 2) στοχεύει στην αξιολόγηση της ικανότητας *Παραγωγής Γραπτού Λόγου* (Writing), αλλά και της *Γραπτής Διαμεσολάβησης*⁷. Η τρίτη επικεντρώνεται στην εξέταση της *Κατανόησης Προφορικού Λόγου* (Listening Comprehension). Τέλος, η Προφορική Εξέταση έχει μία μόνο εξεταστική ενότητα, την *Παραγωγή Προφορικού Λόγου* (Speaking).

Για να επιτύχει κανείς στις εξετάσεις θα πρέπει να συγκεντρώσει το 60% της μέγιστης βαθμολογίας τόσο στην γραπτή, όσο και στην προφορική εξέταση. Επειδή όμως η γραπτή εξέταση αποτελείται από τρεις ξεχωριστές εξεταστικές ενότητες, μια επιπλέον προϋπόθεση επιτυχίας είναι η εξασφάλιση του 30% σε κάθε φάση χωριστά.

4. Τεστ-κριτήρια

Τα τεστ-κριτήρια που θα χρησιμοποιηθούν για την έρευνα της σύγχρονης εγκυρότητας του ΚΠΓ είναι δύο: το QPT και το FCE. Παρά το γεγονός ότι το QPT αποτελεί ένα πλέον αναγνωρισμένο κατατακτήριο τεστ, εφόσον δεν εξετάζει παρά μόνο τις δεξιότητες *Κατανόησης Γραπτού Λόγου* καθώς και *Χρήσης της Γλώσσας*⁸, δεν είναι δυνατή η εξέταση του βαθμού συσχέτισης μία προς μία και των υπόλοιπων δεξιοτήτων που αξιολογούνται από το ΚΠΓ. Γι' αυτό το λόγο, κρίθηκε αναγκαίο να χρησιμοποιηθεί ως τεστ-κριτήριο και το FCE, καθώς εξετάζει τις ίδιες δεξιότητες με το ΚΠΓ, γεγονός που επιτρέπει τον έλεγχο του βαθμού 'σύγκλισης' μεταξύ τους μία προς μία.

Βασικός λόγος επιλογής του FCE ως ένα ακόμα τεστ-κριτήριο είναι το γεγονός ότι πρόκειται για ένα παλαιότερο τεστ επάρκειας, το οποίο, όχι μόνο έχει αναγνωριστεί και καθιερωθεί, αλλά και έχει δημοσιεύσει αρκετές μελέτες πάνω στην εγκυρότητά του, αποτελώντας το πλέον διαδεδομένο στην κατηγορία του. Επίσης, ανήκει στο επίπεδο γλωσσομάθειας B2 του CEFR, όπως και το ΚΠΓ. Μια τέτοια *μελέτη συγκρισιμότητας* (comparability study) του ΚΠΓ με άλλες αναγνωρισμένες εξετάσεις του ίδιου επιπέδου, όπως, για παράδειγμα, το FCE, προτείνεται άλλωστε και από τους Eckes et al. (2005: 362).

4.1 Quick Placement Test (QPT)

Όσον αφορά το QPT, ως ένα αντικειμενικό και *σταθμισμένο τεστ* (standardized test) και δεδομένης της αξιοπιστίας και της εγκυρότητάς του, μπορεί να χρησιμοποιηθεί ως μέτρο σύγκρισης της εγκυρότητας του ΚΠΓ.

⁶ Η Γλωσσική Συνειδητότητα του υποψήφιου επικεντρώνεται στην ενσυνείδητη χρήση της γλώσσας για την επίτευξη επικοινωνιακών στόχων.

⁷ Στο ζητούμενο της *Γραπτής Διαμεσολάβησης*, ο υποψήφιος καλείται να συντάξει ένα γραπτό κείμενο συνεχούς λόγου στην αγγλική γλώσσα με αφορμή κείμενο στην ελληνική.

⁸ Η έκδοση του QPT για υπολογιστή εξετάζει και *Κατανόηση Προφορικού Λόγου*. Ωστόσο, για πρακτικούς λόγους στην παρούσα έρευνα θα χρησιμοποιηθεί η έντυπη έκδοση, η οποία εξετάζει μόνο *Κατανόηση Γραπτού Λόγου* και *Χρήση της Γλώσσας*.

Το QPT είναι ένα σύντομο κατατακτήριο τεστ (Quick Placement Test) επάρκειας της αγγλικής γλώσσας το οποίο εξετάζει την *Κατανόηση Γραπτού Λόγου* (Reading), καθώς και τη *Χρήση της Γλώσσας* (Use of English). Όλα τα ερωτήματα είναι σε μορφή *πολλαπλών επιλογών* (multiple-choice), γεγονός που αυξάνει την αξιοπιστία του.

Το QPT δεν απευθύνεται σε συγκεκριμένο επίπεδο γλωσσομάθειας των εξεταζόμενων. Μπορεί να χρησιμοποιηθεί από υποψήφιους όλων των ηλικιών και επιπέδων γλωσσομάθειας της αγγλικής γλώσσας. Η έντυπη έκδοση ξεκινάει από το *Στοιχειώδες επίπεδο* (Elementary level) με το επίπεδο δυσκολίας να αυξάνει προοδευτικά ως και το *Πολύ Προχωρημένο επίπεδο* (Very Advanced level).

Ένα σημαντικό πλεονέκτημα του QPT είναι το γεγονός ότι η αναφορά των αποτελεσμάτων γίνεται σύμφωνα με την κλίμακα των πέντε επιπέδων γλωσσομάθειας του ALTE (Association of Language Testers in Europe)⁹.

4.2 First Certificate in English (FCE)¹⁰

Το FCE αποτελεί το τρίτο επίπεδο των εξετάσεων του Cambridge ESOL (English for Speakers of Other Languages). Το Cambridge ESOL ανήκει στο UCLES (University of Cambridge Local Examinations Syndicate), το οποίο παρέχει εξετάσεις στην αγγλική γλώσσα από το 1913. Οι εξετάσεις του Cambridge ESOL είναι μέλος του ALTE.

Η εξέταση του FCE καλύπτει τις ακόλουθες *γλωσσικές δεξιότητες* (language skills): *Κατανόηση Γραπτού Λόγου* (paper 1), *Παραγωγή Γραπτού Λόγου* (paper 2), *Χρήση της Αγγλικής* (paper 3), *Κατανόηση Προφορικού Λόγου* (paper 4) και *Παραγωγή Προφορικού Λόγου* (paper 5).

Καθεμία από τις πέντε ενότητες του FCE συμβάλει ισάξια στην τελική βαθμολογία, εφόσον συγκεντρώνει 40 από τους συνολικούς 200 βαθμούς, δηλαδή το 1/5. Ακόμη, σε καμία από τις εξεταστικές ενότητες δεν υπάρχει ελάχιστο ικανοποιητικό όριο για να περάσει κανείς τις εξετάσεις.

Το τελικό αποτέλεσμα δε δίνεται με αριθμούς, αλλά με τη μορφή πέντε ‘περιγραφικών’ βαθμολογιών: A, B, C, D και E. Οι τρεις πρώτες, A, B και C, αντιστοιχούν σε διαφορετικά επίπεδα επιτυχίας, ενώ οι υπόλοιπες δύο, D και E, σε διαφορετικό βαθμό αποτυχίας.

Μετά τις εξετάσεις τα γραπτά αποστέλλονται στο Cambridge για αξιολόγηση. Περίπου δύο μήνες μετά όλοι οι υποψήφιοι –επιτυχόντες και μη– λαμβάνουν μια *κατάσταση αποτελεσμάτων* (statement of results) με τις επιδόσεις τους αναλυτικά σε κάθε εξεταστική ενότητα. Με αυτό τον τρόπο, δίνεται το ‘*προφίλ*’ του κάθε υποψήφιου (candidate profile) μέσα από μια σχηματική περιγραφή της απόδοσής του σε κάθε εξεταστική ενότητα. Η απόδοση του υποψήφιου απεικονίζεται με βάση την ακόλουθη κλίμακα: *Εξαιρετική* (exceptional), *καλή* (good), *οριακή* (borderline), και *αδύνατη* (weak).

Μετά την περιγραφή των παραπάνω τεστ-κριτηρίων, ακολουθούν οι υποθέσεις εγκυρότητας κριτηρίου του ΚΠΓ.

5. Υποθέσεις

Η βασική υπόθεση έρευνας όσον αφορά την εγκυρότητα κριτηρίου του ΚΠΓ είναι η ακόλουθη:

Το ΚΠΓ έχει εγκυρότητα κριτηρίου, και συγκεκριμένα σύγχρονη εγκυρότητα.

⁹ Ο ALTE είναι ένας σύνδεσμος εξεταστών ξένων γλωσσών στην Ευρώπη με σκοπό μεταξύ άλλων την καθιέρωση κοινών επιπέδων γλωσσομάθειας.

¹⁰ Οι πληροφορίες που ακολουθούν βασίζονται στο *Handbook for Teachers* του FCE (2005).

Όπως έχει αναφερθεί, ένα τεστ διαθέτει εγκυρότητα κριτηρίου, όταν διαπιστώνεται σημαντική *θετική συσχέτιση* (positive correlation) μεταξύ των αποτελεσμάτων του και ενός ή και περισσότερων εξωτερικών κριτηρίων.

Άρα, η παραπάνω υπόθεση έχει ως εξής:

Υπάρχει σημαντική θετική συσχέτιση μεταξύ των αποτελεσμάτων στο ΚΠΓ και σε ένα ή και περισσότερα τεστ-κριτήρια.

Η απάντηση στην παραπάνω υπόθεση μπορεί να δοθεί με τη σύγκριση των αποτελεσμάτων των υποψηφίων στο ΚΠΓ και στα δύο τεστ-κριτήρια που χρησιμοποιήθηκαν στην έρευνα με την προϋπόθεση βέβαια ότι όλα τα τεστ διεξάγονται την ίδια χρονική περίοδο.

Οπότε και οδηγούμαστε στις παρακάτω υποϋποθέσεις:

1. *Υπάρχει σημαντική θετική συσχέτιση μεταξύ των αποτελεσμάτων στο ΚΠΓ και στο τεστ-κριτήριο QPT.*
2. *Υπάρχει σημαντική θετική συσχέτιση μεταξύ των αποτελεσμάτων στο ΚΠΓ και στο τεστ-κριτήριο FCE.*

Εφόσον το FCE εξετάζει τις ίδιες δεξιότητες με το ΚΠΓ, είναι δυνατός ο έλεγχος των συσχετίσεων μεταξύ των δύο, όχι μόνο σε επίπεδο τελικού αποτελέσματος, αλλά και σε κάθε εξεταστική ενότητα χωριστά. Επομένως, η παραπάνω υπόθεση αποτελείται από τις ακόλουθες υποϋποθέσεις:

- 2.1 *Υπάρχει σημαντική θετική συσχέτιση μεταξύ των αποτελεσμάτων στο ΚΠΓ και στο FCE σε επίπεδο τελικού αποτελέσματος &*
- 2.2 *Υπάρχει σημαντική θετική συσχέτιση μεταξύ των αποτελεσμάτων στο ΚΠΓ και στο FCE σε επίπεδο κάθε εξεταστικής ενότητας χωριστά.*

6. Έλεγχος σύγχρονης εγκυρότητας του ΚΠΓ

6.1 Μέθοδος

Για την έρευνα της εγκυρότητας κριτηρίου του ΚΠΓ χρησιμοποιήθηκαν τα αποτελέσματα των υποκειμένων στο ΚΠΓ και στα δύο τεστ-κριτήρια που περιγράφηκαν παραπάνω, QPT και FCE, κατά τη χρονική περίοδο 2003-2004. Πιο συγκεκριμένα, χρησιμοποιήθηκαν υποκείμενα από δύο εξεταστικές περιόδους των ΚΠΓ και FCE, Νοεμβρίου-Δεκεμβρίου 2003¹¹ και Μαΐου-Ιουνίου 2004¹².

Η πλειονότητα των υποκειμένων παρακολουθούσε μαθήματα σε φροντιστήρια ξένων γλωσσών¹³. Λίγο πριν τη διεξαγωγή των εξετάσεων, δόθηκε στα υποκείμενα και ένα τρίτο τεστ, το QPT.

Παρόλο που το συγκεκριμένο τεστ παρουσιάστηκε τόσο στους υποψηφίους όσο και στους καθηγητές τους ως ένα τυποποιημένο κατατακτήριο τεστ που θα μπορούσε να 'προβλέψει' ως ένα σημείο και τη μελλοντική απόδοση των μαθητών στα δύο τεστ, ΚΠΓ και FCE, δεν ήταν εύκολο να πειστούν να αφιερώσουν λίγο από το χρόνο τους, ειδικά τη συγκεκριμένη χρονική περίοδο λίγο πριν τις εξετάσεις. Γι' αυτό το λόγο, πολλοί ήταν που δεν έδωσαν το QPT, καθώς δεν έλειψαν και αυτοί που –ενώ τελικά συμμετείχαν– δεν το πήραν στα σοβαρά, με αποτέλεσμα να μειωθεί σημαντικά ο συνολικός αριθμός των υποκειμένων που συμπεριλήφθησαν στην έρευνα, τουλάχιστον όσον αφορά το QPT.

Η συγκεκριμένη χρονική στιγμή διεξαγωγής του τεστ (λίγο πριν τις εξετάσεις) απέτρεψε, ωστόσο, τη μεσολάβηση μεγάλου χρονικού διαστήματος μεταξύ των διαφορετικών τεστ, κατά το οποίο θα μπορούσε να μεταβληθεί η απόδοση των υποκειμένων. Σε

¹¹ Το 2003 οι εξετάσεις του ΚΠΓ διεξήχθησαν το Νοέμβριο, ενώ αντίστοιχα του FCE το Δεκέμβριο.

¹² Το 2004 οι εξετάσεις του ΚΠΓ διεξήχθησαν τον Ιούνιο, ενώ αντίστοιχα του FCE το Μάιο.

¹³ Τα φροντιστήρια που δέχθηκαν να συνεργαστούν βρίσκονται σε διαφορετικές περιοχές της Θεσσαλονίκης (Κέντρο, Ανατολική και Δυτική Θεσσαλονίκη), καθώς και στην Ημαθία, Καβάλα, Κοζάνη, Πτολεμαΐδα, και Αθήνα.

αυτό βοήθησε βέβαια και το γεγονός ότι τα δύο τεστ ΚΠΓ και FCE τόσο το 2003 όσο και το 2004 διεξήχθησαν σε κοντινές χρονικές περιόδους (Νοέμβριο-Δεκέμβριο και Μάιο-Ιούνιο).

Μετά τις εξετάσεις των ΚΠΓ και FCE, συγκεντρώθηκαν τα επίσημα αποτελέσματα των υποκειμένων στα δύο τεστ¹⁴. Όσον αφορά το FCE –όπως έχει αναφερθεί και παραπάνω– η βαθμολογία που αποστέλλεται στους υποψήφιους δεν είναι αριθμητική, αλλά περιγραφική. Για να υπάρξει ωστόσο η δυνατότητα σύγκρισης των αποτελεσμάτων των υποκειμένων στα διαφορετικά τεστ, έγινε μετατροπή της συγκεκριμένης βαθμολογίας σε αριθμητική¹⁵ με βάση τα στοιχεία που δίνονται στην κατάσταση αποτελεσμάτων του FCE.

6.2 Υποκείμενα

Στην έρευνα πήραν μέρος 124 υποκείμενα. Από αυτά, 33 έδωσαν και τα τρία τεστ, 105 έδωσαν τα δύο τεστ, ΚΠΓ και FCE, 49 έδωσαν το ΚΠΓ και το QPT, και 36 έδωσαν το FCE και το QPT.

Όπως αναφέρθηκε και παραπάνω, τα υποκείμενα ήταν υποψήφιοι στις εξετάσεις των ΚΠΓ και FCE κατά τις εξεταστικές περιόδους Νοεμβρίου-Δεκεμβρίου 2003 και Μαΐου-Ιουνίου 2004. 84 (66,1%) από τα υποκείμενα πήραν μέρος στις εξετάσεις της πρώτης περιόδου, και τα υπόλοιπα 42 (33,9%) σε αυτές της δεύτερης εξεταστικής περιόδου.

Από τα ερωτηματολόγια που μοιράστηκαν στα υποκείμενα, προκύπτουν τα εξής: Καταρχήν, όλα τα υποκείμενα ήταν γνώστες της ελληνικής γλώσσας, εφόσον κάτι τέτοιο θεωρείται απαραίτητο για την εξέταση του ΚΠΓ, λόγω των σημείων στην κατανόηση γραπτού και προφορικού λόγου που απαιτούν διαμεσολάβηση.

Η μέση ηλικία (mean) ήταν 17,8. Ο μικρότερος σε ηλικία ήταν 13, ενώ ο μεγαλύτερος 35. Η αναλογία των γυναικών (59,7%) ήταν μεγαλύτερη από αυτή των ανδρών (40,3%). Όσον αφορά το μορφωτικό επίπεδο των υποκειμένων, το 63,7% ήταν μαθητές γυμνασίου-λυκείου, το 22,6% σπουδαστές-φοιτητές, το 4,8% απόφοιτοι λυκείου, και το 8,9% απόφοιτοι τεχνολογικών και πανεπιστημιακών σχολών. Τέλος, το μεγαλύτερο μέρος των εξεταζόμενων είχε παρακολουθήσει προπαρασκευαστικά μαθήματα πριν τις εξετάσεις.

Η κανονικότητα της κατανομής των αποτελεσμάτων των υποκειμένων του δείγματος και στα τρία τεστ επιβεβαιώνεται και με το One Sample Kolmogorov-Smirnov Test (1-Sample K-S Test)¹⁶. Τόσο η τιμή του για το ΚΠΓ (0,200), όσο και για το FCE (0,154) και το QPT (0,200) είναι μεγαλύτερες του 0,05 –γεγονός ενδεικτικό της κανονικής τους κατανομής.

Όσον αφορά το μέσο όρο των επιδόσεων του δείγματος τόσο στο ΚΠΓ (65,30), όσο και στο FCE (65,49), είναι υψηλότερος από τους αντίστοιχους του πληθυσμού, γεγονός που οφείλεται στο ότι πολλά υποκείμενα, ενώ αρχικά πήραν μέρος στην έρευνα, στη συνέχεια απέκριναν τα αποτελέσματά τους στα ΚΠΓ και FCE λόγω της αποτυχίας τους.

¹⁴ Τα δεδομένα του ΚΠΓ μου παραχωρήθηκαν από τον πρόεδρο της εξεταστικής επιτροπής του ΚΠΓ, Α. Τσομπάνογλου.

¹⁵ Όπως βέβαια αναφέρεται και στην κατάσταση αποτελεσμάτων του FCE, το σύνολο της βαθμολογίας που χρειάζεται να συγκεντρώσει κανείς για να του αποδοθεί Α, Β, C, D ή Ε μπορεί ελαφρώς να ποικίλλει, ανάλογα με τη δυσκολία των θεμάτων της κάθε εξεταστικής περιόδου. Γι' αυτό και η μετατροπή των περιγραφικών βαθμολογιών σε αριθμητικές είναι κατά προσέγγιση.

¹⁶ Το 1-Sample K-S Test χρησιμοποιείται για τον έλεγχο κανονικής κατανομής των αποτελεσμάτων σε ένα τεστ. Οι συγκεκριμένες τιμές που αναφέρονται εδώ προκύπτουν με τη διόρθωση "Lilliefors Significance Correction". Βλ. και SPSS 11.0.

Ωστόσο, η αναλογία της επιτυχίας στα ΚΠΓ και FCE (0,97) του δείγματος είναι αντίστοιχη της αναλογίας της επιτυχίας στα ΚΠΓ και FCE του πληθυσμού, τόσο για το 2003 (0,98) όσο και για το 2004 (0,87).

Επίσης, αξίζει να σημειωθεί ότι ο συντελεστής ομοιογένειας¹⁷ του δείγματος, όσον αφορά τα αποτελέσματα στο ΚΠΓ¹⁸ (16,8%), εμφανίζεται αρκετά υψηλός και μάλιστα μεγαλύτερος του αντίστοιχου συντελεστή του πληθυσμού, τόσο για το ΚΠΓ της εξεταστικής του 2003 (23,3%) όσο και για το ΚΠΓ της εξεταστικής του 2004 (23,1%).

Στον ακόλουθο πίνακα περιγράφονται οι μέσοι όροι των υποκειμένων στα ΚΠΓ και FCE. Είναι φανερό ότι οι μέσοι όροι τόσο στο τελικό αποτέλεσμα όσο και στις επιμέρους εξεταστικές ενότητες¹⁹ είναι πολύ ‘κοντά’. Ειδικά όσον αφορά το τελικό αποτέλεσμα αλλά και τα αποτελέσματα στις δύο τελευταίες ενότητες κατανόησης και παραγωγής προφορικού λόγου, οι μέσοι όροι είναι σχεδόν ίσοι.

Πίνακας 1. Μέσοι όροι των επιδόσεων του δείγματος στα ΚΠΓ & FCE

	Συχνότητα	Ελάχιστο	Μέγιστο	Μέση τιμή	Τυπική Απόκλιση
ΚΠΓ	124	38	89	65,30	10,98
FCE	108	44	85	65,49	8,65
ΚΠΓ1	124	32	92	69,36	12,74
FCE1 ²⁰	88	43	84	66,10	9,16
ΚΠΓ2	124	15	98	60,17	16,92
FCE2	88	44	88	65,95	10,20
ΚΠΓ3	124	20	100	64,26	14,35
FCE3	88	44	84	64,47	10,17
ΚΠΓ4	124	30	100	68,31	17,42
FCE4	88	44	84	68,26	9,69

7. Αποτελέσματα

7.1 Συσχέτιση μεταξύ του ΚΠΓ και των τεστ-κριτηρίων

Τα αποτελέσματα των υποκειμένων κωδικοποιήθηκαν και αναλύθηκαν στατιστικά μέσω του στατιστικού προγράμματος SPSS²¹ 11.0. Ο δείκτης που χρησιμοποιήθηκε για τον έλεγχο των συσχετίσεων μεταξύ του ΚΠΓ και των τεστ-κριτηρίων είναι ο δείκτης

¹⁷ Ο συντελεστής ομοιογένειας υπολογίζεται από την αναλογία της τυπικής απόκλισης προς το μέσο όρο.

¹⁸ Ο συντελεστής ομοιογένειας του πληθυσμού του FCE δε μπορεί να υπολογιστεί, εφόσον δε γνωρίζουμε την τυπική απόκλιση.

¹⁹ Επειδή στο ΚΠΓ δε συμβάλλουν όλες οι εξεταστικές ενότητες στον ίδιο βαθμό στην τελική βαθμολογία, έγινε μετατροπή των βαθμολογιών σε εκατοστιαία κλίμακα για να υπάρξει η δυνατότητα σύγκρισης με τις επιμέρους βαθμολογίες στο FCE.

²⁰ Ως FCE1 –όπως θα δούμε αμέσως παρακάτω– συμβολίζεται ο μέσος όρος της βαθμολογίας των υποκειμένων στις εξεταστικές ενότητες FCE1 & FCE3. Αυτό γίνεται για να υπάρχει αντιστοιχία με την εξεταστική ενότητα του ΚΠΓ (ΚΠΓ1), στην οποία εξετάζονται μαζί οι δεξιότητες που εξετάζονται στα FCE1 & FCE3.

²¹ Statistical Package for the Social Sciences.

Pearson (Pearson correlation coefficient), ο οποίος και χρησιμοποιείται περισσότερο στις επιστημονικές έρευνες των κοινωνικών επιστημών.

Ο χαρακτηρισμός της σχέσης μεταξύ δύο μεταβλητών ως ισχυρή, μέτρια, ή αδύνατη δεν είναι προκαθορισμένη. Ωστόσο, σύμφωνα και με τον Κατσίλλη (2000: 127), μεταξύ άλλων, μια σχέση με συντελεστή συσχέτισης από 0 ως 0,3 θεωρείται αδύνατη, από 0,3 ως 0,6 μέτρια και από 0,6 ως 1 ισχυρή²².

Το QPT εξετάζει τις δεξιότητες *Κατανόησης Γραπτού Λόγου και Χρήσης της Γλώσσας*. Οι ίδιες δεξιότητες ελέγχονται στο πρώτο θέμα εξέτασης του ΚΠΓ (ΚΠΓ1). Ωστόσο, ο έλεγχος εγκυρότητας κριτηρίου του ΚΠΓ με το QPT δεν περιορίζεται στο ΚΠΓ1, αλλά υπολογίζεται το τελικό αποτέλεσμα στο ΚΠΓ. Άλλωστε, το ΚΠΓ1 εμφανίζει υψηλή συσχέτιση (0,748) με το τελικό αποτέλεσμα στο ΚΠΓ. Εξάλλου, το QPT, παρόλο που εξετάζει μόνο τις δεξιότητες *Κατανόησης Γραπτού Λόγου και Χρήσης της Γλώσσας*, χρησιμοποιείται ευρέως για την κατάταξη υποκειμένων σε επίπεδα γλωσσομάθειας της Αγγλικής.

Όπως φαίνεται και στον ακόλουθο πίνακα, οι δείκτες συσχέτισης Pearson μεταξύ του ΚΠΓ και των δύο τεστ-κριτηρίων, QPT (0,694) και FCE (0,777), εμφανίζουν σημαντική θετική συσχέτιση²³ σε επίπεδο σημαντικότητας 0,01 επιβεβαιώνοντας έτσι την εγκυρότητά του. Όταν λέμε ότι μία συσχέτιση είναι σημαντική σε επίπεδο 0,01 ($p < 0,01$), αυτό σημαίνει ότι υπάρχει μόνο 1% πιθανότητα η συσχέτιση να οφείλεται στην τύχη. Είμαστε, δηλαδή, σύμφωνα και με τους Bachman (2004: 227) και Brown (1996: 165-166), 99% σίγουροι ότι η συγκεκριμένη τιμή είναι πραγματική.

Ενδιαφέρον αποτελεί το γεγονός ότι και ο δείκτης συσχέτισης ανάμεσα στα QPT και FCE εμφανίζεται στα ίδια υψηλά επίπεδα (0,752). Μάλιστα, το γεγονός ότι το FCE εμφανίζει υψηλή συσχέτιση με το QPT, το οποίο –όπως έχει αναφερθεί– αποτελεί ένα αντικειμενικό και σταθμισμένο τεστ, ενισχύει ακόμη περισσότερο την επιλογή του FCE ως ένα ακόμη τεστ-κριτήριο στην παρούσα έρευνα.

7.2 Συσχέτιση μεταξύ των εξεταστικών ενοτήτων στα ΚΠΓ και FCE

Όσον αφορά τον έλεγχο των συσχετίσεων μεταξύ των εξεταστικών ενοτήτων του ΚΠΓ και του FCE, ισχύουν τα παρακάτω: Όπως έχει ήδη αναφερθεί, στο FCE η *Κατανόηση Γραπτού Λόγου* (Reading Comprehension) και η *Χρήση της Γλώσσας* (Use of English) εξετάζονται χωριστά, στην πρώτη (FCE1) και στην τρίτη (FCE3), αντίστοιχα, εξεταστική ενότητα. Αντίθετα, στο ΚΠΓ οι παραπάνω δεξιότητες εξετάζονται μαζί στην πρώτη εξεταστική ενότητα (ΚΠΓ1) *Κατανόησης Γραπτού Λόγου και Γλωσσικής Συνειδητότητας* (Reading Comprehension & Language Awareness).

Για να υπάρξει, ωστόσο, αντιστοιχία των δεξιοτήτων μία προς μία, οι δύο παραπάνω εξεταστικές ενότητες του FCE (FCE1 & FCE3) ‘συνοψίζονται’ σε μία, η οποία θα ονομάζεται FCE1 από εδώ και στο εξής. Ούτως ή άλλως, το FCE1 εμφανίζει ισχυρή συσχέτιση με το FCE3 (0,629), η οποία μάλιστα είναι και η υψηλότερη *εσωτερική συσχέτιση* (internal correlation) του εν λόγω τεστ. Επίσης, τόσο το FCE1 (0,813) όσο και το FCE3 (0,816), εμφανίζουν την ίδια υψηλή συσχέτιση με το τελικό αποτέλεσμα στο

²² Ωστόσο, κάτι τέτοιο δεν είναι απόλυτο καθώς η ταξινόμηση μιας σχέσης μεταξύ μεταβλητών βασίζεται και στο εκάστοτε αντικείμενο μελέτης, όπως και σε προηγούμενες μελέτες πάνω σε αυτό.

²³ Στην ερμηνεία των δεικτών συσχέτισης, θα πρέπει να είναι μεγαλύτερη από την *κριτική συσχέτιση* (critical correlation), για να θεωρηθεί *στατιστικά σημαντική* (statistically significant). Οι κριτικές τιμές του δείκτη συσχέτισης Pearson καθορίζονται με βάση τον αριθμό των υποκειμένων N. Όσο μικρότερο είναι το N, τόσο μεγαλύτερη είναι η κριτική τιμή, και αντίστροφα. Ωστόσο, το SPSS ελέγχει αυτόματα τη σημαντικότητα των δεικτών συσχέτισης με βάση την αντίστοιχη κριτική τιμή κάθε φορά. Βλ. Bachman (2004: 251-253) και Brown (1996: 162-166).

FCE, γεγονός που υποδηλώνει την ισοδύναμη ‘συμβολή’ τους στο συνολικό αποτέλεσμα.

Κατά συνέπεια, και για να μην υπάρξει σύγχυση με τη σειρά των εξεταστικών ενοτήτων, θα ακολουθήσουμε και στο FCE την ίδια αρίθμηση με το ΚΠΓ. Δηλαδή, θα ισχύουν τα εξής:

ΚΠΓ1 – FCE1: Κατανόηση Γραπτού Λόγου & Γλωσσική Συνειδητότητα
(Reading Comprehension & Language Awareness)

ΚΠΓ2 – FCE2: Παραγωγή Γραπτού Λόγου (Writing)

ΚΠΓ3 – FCE3: Κατανόηση Προφορικού Λόγου (Listening Comprehension)

ΚΠΓ4 – FCE4: Παραγωγή Προφορικού Λόγου (Speaking)

Σε όλες τις εξεταστικές ενότητες των δύο τεστ παρατηρείται σημαντική θετική συσχέτιση σε επίπεδο σημαντικότητας 0,01. Την υψηλότερη θετική συσχέτιση εμφανίζουν οι εξεταστικές ενότητες ΚΠΓ1 & FCE1 (0,617), οι οποίες ελέγχουν τη δεξιότητα *Κατανόησης Γραπτού Λόγου*, καθώς και αυτήν της *Γλωσσικής Συνειδητότητας*. Ακολουθεί η συσχέτιση μεταξύ των ΚΠΓ2 & FCE2 (0,518), οι οποίες εξετάζουν τη δεξιότητα *Παραγωγής Γραπτού Λόγου*. Ο έλεγχος *Κατανόησης Προφορικού Λόγου* στα ΚΠΓ3 & FCE3 εμφανίζει μια επίσης σημαντική θετική συσχέτιση (0,397). Τέλος, ο δείκτης συσχέτισης μεταξύ των εξεταστικών ενοτήτων *Παραγωγής Προφορικού Λόγου*, ΚΠΓ4 & FCE4 (0,376) κυμαίνεται στα ίδια επίπεδα με τον προηγούμενο.

8. Ερμηνεία των αποτελεσμάτων

Τα παραπάνω αποτελέσματα αποτελούν μια ουσιαστική απόδειξη της εγκυρότητας κριτηρίου –και συγκεκριμένα της σύγχρονης εγκυρότητας– του ΚΠΓ και με τα δύο τεστ-κριτήρια που χρησιμοποιήθηκαν, QPT και FCE. Το τελικό αποτέλεσμα στο ΚΠΓ εμφανίζει όχι μόνο σημαντική θετική συσχέτιση σε επίπεδο σημαντικότητας 0,01 αλλά και αρκετά υψηλή, τόσο με το QPT, όσο και με το FCE, γεγονός που οδηγεί στην επιβεβαίωση των δύο υποϋποθέσεων της έρευνας. Ακόμη, οι σημαντικές θετικές συσχετίσεις μεταξύ των αποτελεσμάτων στις διαφορετικές εξεταστικές ενότητες των ΚΠΓ και FCE αποτελούν μία ακόμη επιβεβαίωση της εγκυρότητας κριτηρίου του ΚΠΓ.

Ωστόσο, σύμφωνα και με τον Messick (1995: 746), η συγκεκριμένη μορφή εγκυρότητας θα πρέπει να στοχεύει όχι μόνο στην απόδειξη *σύγκλισης* (convergence) των αποτελεσμάτων του τεστ με συγκεκριμένα εξωτερικά κριτήρια που μετρούν το ίδιο ‘construct’, αλλά και στην απόδειξη *απόκλισης* τους (divergence) από άλλα κριτήρια άσχετα με το ‘construct’ του τεστ υπό αξιολόγηση. Πρόταση για περαιτέρω έρευνα, όσον αφορά τη διερεύνηση της σύγχρονης εγκυρότητας του ΚΠΓ, θα μπορούσε λοιπόν να αποτελέσει ο ταυτόχρονος έλεγχος τόσο της *συγκλίνουσας* (convergent validity) όσο και της *αποκλίνουσας εγκυρότητάς* του (divergent validity), κάτι το οποίο αποτελεί άλωστε απαραίτητο ‘συστατικό’ του συνολικού ελέγχου της εγκυρότητας ενός τεστ²⁴.

Τέλος, δε θα πρέπει να ξεχνάμε ότι ο έλεγχος εγκυρότητας ενός τεστ θα πρέπει να αποτελεί μια συνεχή και συλλογική διαδικασία, η οποία ξεκινάει με τη ‘γέννησή’ του και θα πρέπει να συνεχίζεται σε όλη τη διάρκεια της ύπαρξής του, παρά τα πιθανά πρακτικά προβλήματα που μπορεί να ‘συνοδεύουν’ μια τέτοιου είδους πολύπλευρη έρευνα.

²⁴ Μια ιδιαίτερα σημαντική μέθοδος ελέγχου των παραπάνω μορφών εγκυρότητας αποτελεί και το *Σχέδιο Πολλαπλών Ικανοτήτων-Πολλαπλών Μεθόδων* (Multitrait-Multimethod matrix) ή MTMM. Για την εφαρμογή της μεθόδου MTMM βλ. και Campbell & Fiske (1959), Boldt, & Oltman (1993), Bachman, & Palmer (1981), Stevenson (1981), Geranpayeh (1999), και Τσομπάνογλου (2008, 2009).

Βιβλιογραφία

- Association of Language Testers in Europe. (1998). *A Multilingual Glossary of Language Testing Terms*. Cambridge: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F. & A. S. Palmer. (1981). A multitrait-multimethod investigation into the construct validity of six tests of speaking and reading. In A. S. Palmer, P. J. M. Groot & G. A. Trosper (eds.) *The Construct Validation of Tests of Communicative Competence*. Washington: TESOL, 149-65.
- Boldt, R. F. & P. K. Oltman. (1993). *Multimethod construct validation of the Test of Spoken English. TOEFL Research Report 46*. Princeton, N.J.: Educational Testing Service.
- Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River, N.J.: Prentice Hall Regents.
- Brown, J. D. (2000). "What is construct validity?" In *Shiken: JALT Testing and Evaluation SIG Newsletter* 4/2. <http://www.jalt.org/test/bro_8.htm>
- Eckes T. et al. (2005). "Progress and problems in reforming public language examinations in Europe: cameos from the Baltic States, Greece, Hungary, Poland, Slovenia, France and Germany". *Language Testing* 22(3): 355-377.
- Geranpayeh, A. (1999). *Language proficiency testing: A comparative analysis of IELTS and TOEFL*. Unpublished PhD thesis, University of Edinburgh.
- Hatch, E. & H. Farhady. (1982). *Research Design and Statistics for Applied Linguistics*. Rowley, Mass.: Newbury House Publishers, INC.
- Heaton, J. B. (1975). *Writing English Language Tests: A Practical Guide for Teachers of English as a Second or Foreign Language*. London: Longman.
- Κατσύλλης, Ι. Μ. (2000). *Περιγραφική Στατιστική Εφαρμοσμένη στις Κοινωνικές Επιστήμες και την Εκπαίδευση*. Αθήνα: Gutenberg.
- Messick, S. (1995). "Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning". *American Psychologist* 50(9): 741-749.
- Palmer, A. S., P. J. M. Groot & G. A. Trosper (eds.). (1981). *The Construct Validation of Tests of Communicative Competence*. Washington: TESOL.
- Stevenson, D. K. (1981). Beyond faith and face validity: the multitrait-multimethod matrix and the convergent and discriminant validity of oral proficiency tests; Incl. proc. of colloquium at TESOL '79. In A. S. Palmer, P. J. M. Groot & G. A. Trosper (eds.) *The Construct Validation of Tests of Communicative Competence*. Washington: TESOL, 37-66.
- Τσομπάνογλου, Μ. (2008). *Πρότυπο Αξιολόγησης της Εγκυρότητας Δομής Συστημάτων Πιστοποίησης Γλωσσομάθειας: Εφαρμογή του στο Κρατικό Πιστοποιητικό Γλωσσομάθειας του Επιπέδου B2 της Αγγλικής Γλώσσας*. Διδακτορική διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.
- Τσομπάνογλου, Μ. (2009). "Πρότυπο αξιολόγησης της εγκυρότητας δομής συστημάτων πιστοποίησης γλωσσομάθειας". Στο *Ζητήματα Πιστοποίησης της Γλωσσομάθειας. Το Κρατικό Πιστοποιητικό Γλωσσομάθειας ως Σημείο Αναφοράς*. Θεσσαλονίκη: Μαλλιάρης-Παιδεία, 9-44.
- University of Cambridge ESOL Examinations. (2005). *First Certificate in English: Handbook for Teachers*. Cambridge: UCLES.
- Weir, C. J. (1988). *Communicative Language Testing with Special Reference to English as a Foreign Language*. Exeter: University of Exeter.
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. New York: Palgrave Macmillan.