

Ανάλυση κατά Rasch ενός συνόλου ερωτημάτων αστρονομίας

Αθανάσιος Βέρδης, Σπύρος Κόκκοτας, Ευαγγελία Πάνου

Περίληψη

Η παρούσα εισήγηση εξετάζει πώς η ανάλυση κατά Rasch και η Ανάλυση Κύριων Συνιστωσών μπορούν να χρησιμοποιηθούν συμπληρωματικά η μία προς την άλλη για την αξιολόγηση μια διδακτικής παρέμβασης στο αντικείμενο της αστρονομίας. Για τον σκοπό αυτόν δώδεκα διχοτομικές ερωτήσεις του τύπου «ναι» ή «όχι» απαντήθηκαν από 134 μαθητές και μαθήτριες της τρίτης γυμνασίου σε σχολεία του Πειραιά, της Αθήνας και της Χίου. Το αντικείμενο των ερωτήσεων ήταν ο βαθμός εκπλήρωσης των στόχων του προγράμματος STEMforYouth, το οποίο συμπεριελάμβανε την διδασκαλία εννοιών Αστρονομίας. Η μη γραμμική Ανάλυση Κυρίων Συνιστωσών και η ανάλυση κατά Rasch, συγκρίνονται ως προς το ζήτημα του ελέγχου της μονοδιαστατικότητας στις απαντήσεις των συμμετεχόντων. Υποστηρίζεται ότι ενώ η μη γραμμική Ανάλυση Κυρίων Συνιστωσών προσφέρει ασφαλέστερα κριτήρια για τον έλεγχο της μονοδιαστατικότητας, η ανάλυση κατά Rasch, όταν η εν λόγω μονοδιαστατικότητα γίνει αποδεκτή, προσφέρει περισσότερη πληροφορία. **Λέξεις κλειδιά:** Αστρονομία, STEM, Ανάλυση κατά Rasch.

Rasch analysis of a collection of Astronomy items

Abstract

This paper examines how Rasch analysis and Principal Components Analysis have been used complementarily in the evaluation of an astronomy course. Twelve dichotomous items were presented to 134 secondary school students in Greece. The items evaluated a STEM project, with emphasis on its Astronomy component. Nonlinear Principal Components analysis and Rasch analysis are compared regarding the issue of unidimensionality. It is argued that non-linear Principal Component Analysis offers safer criteria for suggesting unidimensionality but at the same time Rasch analysis provides more information regarding items' difficulty and participants' ability. **Keywords:** Astronomy, STEM, Rasch analysis.

Εισαγωγή

Σκοπός της παρούσας εργασίας είναι να εξετάσει πώς η ανάλυση κατά Rasch και η Ανάλυση Κυρίων Συνιστωσών προσεγγίζουν το ζήτημα της μονοδιαστατικότητας σε ένα σύνολο ερωτημάτων. Η έννοια της «διαστατικότητας» είναι σύνθετη στην ψυχομετρία. Γενικά, θεωρούμε ότι οι αποκρίσεις των εξεταζομένων είναι αναπαραστάσεις κρυφών διαστάσεων. Οι διαστάσεις δεν μετρούνται απευθείας, αλλά είναι αποτέλεσμα συσχετίσεων μεταξύ των ερωτημάτων και των σχετικών απαντήσεων (Gessaroli & de Champlain, 2005). Μονοδιαστατικότητα στη διαδικασία αυτή σημαίνει ότι αν από την συνολική διακύμανση σε σύνολο ερωτημάτων αφαιρέσουμε την διακύμανση που οφείλεται σε μία και μόνη υποτιθέμενη διάσταση, τότε δεν μένει σημαντική διακύμανση στα υπόλοιπα (residuals) της ανάλυσης. Μια κάποια δομή στην διακύμανση των residuals θα οφειλόταν ενδεχομένως σε μία δεύτερη διάσταση.

Το θέμα του αριθμού των διαστάσεων είναι σημαντικό στην εκπαιδευτική έρευνα. Από τις αρχές της δεκαετίας του 1970 μέχρι και τις μέρες μας ο Harvey Goldstein, αρχικά από το London Institute of Education και τελευταία στο Πανεπιστήμιο του Bristol, έχει φέρει στην επιφάνεια το πρόβλημα της εξασφάλιση της μονοδιαστατικότητας στις διεθνείς μελέτες για την αξιολόγηση της μαθητικής επίδοσης τύπου (βλ. Goldstein & Blinkhorn, 1982, Goldstein, 1979a, 1979b). Ο Goldstein και οι συνεργάτες του (βλ. Goldstein, 2004; Kreiner & Christensen, 2013) έχουν ασκήσει κριτική στην εγκυρότητα των διεθνών αξιολογήσεων, οι οποίες βασίζονται στην ανάλυση κατά Rasch και στην νέα ψυχομετρική θεωρία (Item Response Theory). Η συζήτηση για τις στατιστικές προϋποθέσεις της Rasch στην ψυχολογική και εκπαιδευτική έρευνα συνεχίζεται μέχρι σήμερα (βλ. Goldstein, 2010; Panayides, Robinson, & Tymms, 2015), ενώ ένα καλό εισαγωγικό κείμενο στην σχετική συζήτηση είναι αυτό του Wiberg's (2012).

Τα δεδομένα που θα χρησιμοποιήσουμε στην παρούσα έρευνα αφορούν την αξιολόγηση ενός προγράμματος για την διδασκαλία της αστρονομίας. Πρόκειται ουσιαστικά για μια σειρά δέκα μαθημάτων, τα οποία σχεδιάστηκαν από την ελληνική μη-κερδοσκοπική εταιρεία *Research Paths* και πραγματοποιήθηκαν εθελοντικά από εκπαιδευτικούς και μαθητές και δημοσίων σχολείων στην Χίο και στην Αθήνα με την βοήθεια του σχολικού συμβούλου εκπαιδευτικών Φυσικής Κώστα Στεφανίδη. Τα

περιεχόμενα των μαθημάτων ήταν: (α) οι θεωρίες για τον Κόσμο, (β) το ηλιακό σύστημα, (γ) τα ταξίδια στο διάστημα, (δ) ο Ήλιος, (ε) το φαινόμενο του θερμοκηπίου και η κλιματική αλλαγή, (ζ-η) οι τεχνητοί δορυφόροι, (θ) η συμβολή της αστρονομίας στην ανάπτυξη των επιστημών, (ι) η επιστημονική ερευνητική μέθοδος, και (κ) προσομοίωση φαινομένων το σύμπαντος.

Η εννοιολογική κατασκευή της παρούσας έρευνας

Η εννοιολογική κατασκευή της παρούσας έρευνας είναι η δεκτικότητα προς την αστρονομία. Οι Pasachoff και Percy (2005) έχουν συζητήσει τα οφέλη από την διδασκαλία του αντικειμένου αυτού στην σχολική τάξη, ενώ σύγχρονα βιβλία διδακτικής της αστρονομίας συζητούν το «πώς», το «πού», το «γιατί» και το «από ποιους» αυτού του αντικειμένου (Beet, 2015). Τα μαθήματα του εν λόγω μαθήματος όμως εντάσσονταν σε ένα ευρωπαϊκό πρόγραμμα με τίτλο STEM4YOUTH, σκοπός του οποίου ήταν να προωθήσει την δεκτικότητα προς τις θετικές επιστήμες συνολικότερα και όχι μόνο προς την αστρονομία. Στο πλαίσιο αυτό, σκοπός της αξιολόγησης δεν ήταν η μέτρηση του κατά πόσο αυξήθηκαν οι γνώσεις στα διδακτικά αντικείμενα που προσφέρθηκαν, αλλά η μέτρηση του πόσο δεκτικοί έγιναν οι συμμετέχοντες προς τα γνωστικά αντικείμενα του STEM. Πόσο τους αρέσουν τα αντικείμενα αυτά και πόσο πιθανό είναι να ακολουθήσουν ένα επάγγελμα στις θετικές επιστήμες στο μέλλον.

Πέρα από την ελληνική *Research Paths*, στο ευρωπαϊκό αυτό πρόγραμμα συμμετείχαν: (α) το Πολυτεχνείο της Βαρσοβίας για την διδασκαλία εννοιών Φυσικής, (β) το Ευγενίδειο Ίδρυμα για την διδασκαλία της μηχανολογίας, το Ινστιτούτο Τζόζεφ Στέφαν (Σλοβενία) για την διδασκαλία της Χημείας, το Ίδρυμα Ουμπέρτο Βερονέζι (Ιταλία) για την διδασκαλία εννοιών ιατρικής, το Πολυτεχνείο της Οστράβα (Τσεχία) για την διδασκαλία των μαθηματικών, το Πανεπιστήμιο της Βαρκελώνης για την διδασκαλία των φυσικών επιστημών στο κοινό εκτός εκπαίδευσης (*peoples' science*), το Πανεπιστήμιο της Καντάμπρια (Ισπανία), για την αξιολόγηση του προγράμματος, η μη κερδοσκοπική εταιρεία *Open Evidence* (Ισπανία) για την ανάλυση των τάσεων σε σχέση με τα επαγγέλματα STEM στην Ευρώπη και επίσης το Ινστιτούτο Αναπτυξιακών και Στρατηγικών Αναλύσεων (Σλοβενία) για την σύνδεση του προγράμματος με τα κέντρα χάραξης εκπαιδευτικής πολιτικής.

Η αξιολόγηση του STEMforYOUTH ανετέθη στην επιστημονική ομάδα του Πανεπιστημίου της Κατάμπρια, η οποία σχεδίασε μία αξιολόγηση του τύπου «πριν και μετά», βασισμένη στην μέτρηση των απόψεων των συμμετεχόντων μαθητών και εκπαιδευτικών μέσω ερωτηματολογίων. Κάθε ομάδα, συμπεριλαμβανομένης της *Research Paths*, έλαβε σε ηλεκτρονική μορφή τα ερωτηματολόγια, αναλαμβάνοντας ταυτοχρόνως την ευθύνη για την εκτύπωσή τους, καθώς και για την συλλογή, την καταγραφή και την αποστολή των σχετικών δεδομένων στο Πανεπιστήμιο της Κατάμπρια. Οι συμμετέχοντες μαθητές και εκπαιδευτικοί σε κάθε χώρα κλήθηκαν να λάβουν μέρος σε συνεντεύξεις με κοινά ερωτήματα για όλες τις χώρες. Πέρα από την επίσημη διαδικασία αξιολόγησης, η *Research Paths* συμπεριέλαβε και 12 ερωτήματα, καθένα εκ των οποίων θα μπορούσε να απαντηθεί από τους συμμετέχοντες με ένα «ναι» ή ένα «όχι». Τα 12 ερωτήματα παρουσιάζονται, μαζί με μια πρώτη ανάλυση της δυσκολίας τους, στην επόμενη ενότητα.

Η μέτρηση κατά Rasch

Οι απαντήσεις στα 12 ερωτήματα της έρευνας αναλύθηκαν αρχικά με τη βοήθεια ενός μοντέλου Rasch. Η μέτρηση κατά Rasch βασίζεται στην ιδέα του Δανού μαθηματικού Georg Rasch (1980) και μπορεί να εφαρμοστεί σε διχοτομικές αποκρίσεις του τύπου 0 και 1. Οι αποκρίσεις αυτές μπορεί να σημαίνουν «όχι ή ναι» όπως στη δική μας περίπτωση. Σύμφωνα με τον τρόπο του Rasch, η πιθανότητα που έχει ένας εξεταζόμενος k να απαντήσει «ναι» σε ένα ερώτημα l είναι p_{kl} και εξαρτάται από την «ικανότητά» του θ_k , και από την «δυσκολία» του ερωτήματος δ_l . Αν η ικανότητα είναι ίση με την δυσκολία, τότε θεωρούμε ότι ο εξεταζόμενος έχει πιθανότητα να απαντήσει ίση με 0,5. Επειδή η πιθανότητα βρίσκεται στο κλειστό διάστημα 0 ως 1, χρησιμοποιούμε στην Rasch τις μονάδες *logit*, δηλαδή τον λογάριθμο του λόγου των πιθανοτήτων της «σωστής» προς την «λανθασμένη» απάντηση. Έτσι γράφουμε ότι $\log(p_{ij}/1 - p_{ij}) = \theta_i + \delta_j$ και συνεπάγουμε ότι η πιθανότητα να απαντήσει κάποιος «ναι» σε μία ερώτηση είναι:

$$\frac{\exp(\theta_{ij} - \delta_j)}{1 + \exp(\theta_{ij} - \delta_j)}$$

Στον πίνακα που ακολουθεί παρουσιάζουμε τις ερωτήσεις, τις θετικές απαντήσεις («ναι»), την εκτιμώμενη κατά Rasch «δυσκολία» του κάθε ερωτήματος, καθώς και δύο δείκτες καλής προσαρμογής των δεδομένων, στους οποίους θα αναφερθούμε σε επόμενη παράγραφο.

Πίνακας 1. Η χρήση της Rasch και η δυσκολία των προτάσεων σε μονάδες logit.

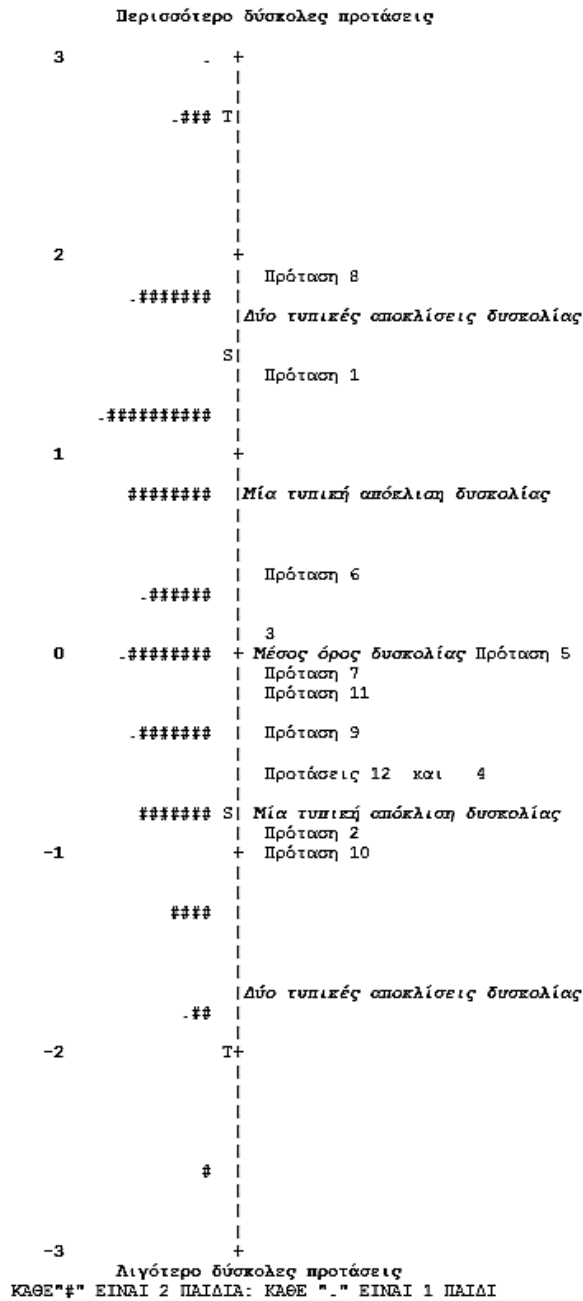
	Ερωτήματα	Θετικές απαντήσεις	Δυσκολία	Τυπικό σφάλμα δυσκολίας	Infit (ZSTD)	Outfit (ZSTD)
8	Έχω διαβάσει βιβλία σε σχέση με το πλανητικό μας σύστημα	31	1,91	0,23	-0,5	0,2
1	Θα ήθελα να ασχοληθώ επαγγελματικά στο μέλλον με τις επιστημονικές κατασκευές	42	1,38	0,21	2,1	2,0
6	Μπορώ να αναγνωρίσω τους αστερισμούς στον νυχτερινό ουρανό	67	0,37	0,20	-1,6	-1,9
3	Μπορώ να φτιάξω ένα τηλεσκόπιο με απλά υλικά στο σπίτι	73	0,13	0,20	1,4	1,4
5	Χρησιμοποιώ τις γνώσεις μου στην αστρονομία για να εξηγήσω διάφορα κοσμικά φαινόμενα	76	0,02	0,20	1,4	1,6
7	Θα ήθελα να εργαστώ ως επιστήμονας ή μηχανικός	80	-0,14	0,20	-0,1	-0,3
11	Στο σπίτι κάνω τεχνολογικές κατασκευές που βλέπω στο διαδίκτυο	82	-0,22	0,20	-0,70	-0,5
9	Μπορώ να γράψω τρεις επιπτώσεις της υπερθέρμανσης του πλανήτη	86	-0,39	0,20	0,3	-0,4
4	Μου αρέσει να φτιάχνω πράγματα στο σπίτι που σχετίζονται με την τεχνολογία	91	-0,60	0,21	-0,6	-0,7
12	Γνωρίζω τέσσερα διαφορετικά επαγγέλματα που συνδέονται με την NASA	92	-0,64	0,21	-0,2	-0,6
2	Αναζητώ επιστημονικές γνώσεις στο διαδίκτυο	97	-0,87	0,22	0,0	-0,4
10	Έχω επισκεφθεί πλανητάριο	99	-0,96	0,22	-1,7	-1,6

Σημαντικό στοιχείο της ανάλυσης που περιγράψαμε πιο πάνω είναι ότι μπορούμε να κατατάξουμε τις ερωτήσεις ενός τεστ ανάλογα με τον βαθμό δυσκολίας της καθεμίας, καθώς και να διακρίνουμε ποιες ερωτήσεις και ποιοι εξεταζόμενοι δεν ταιριάζουν καλά στη λογική της Rasch. Η λογική αυτή είναι ότι οι δυσκολότερες ερωτήσεις θα «νικήσουν» τους λιγότερο ικανούς εξεταζόμενους (Törmäkangas, 2011). Οι προτάσεις αυτές παρουσιάζονται στον Πί-

νακα 1 κατά σειρά δυσκολίας. Για τον υπολογισμό χρησιμοποιήσαμε το λογισμικό *Winsteps*, λόγω της ευκολίας του. Παρόμοια λογισμικά είναι το πακέτο *Itm* στην

γλώσσα προγραμματισμού *R*, καθώς και τα λογισμικά *RUMM*, *Param* και *Rasch*.

Το Σχήμα 1 στα αριστερά, παρουσιάζει στην δεξιά του πλευρά τις προτάσεις από τις πιο δύσκολες (πάνω) ως στις πιο εύκολες (κάτω) και στην αριστερή πλευρά τον αριθμό των συμμετεχόντων. Προτάσεις και συμμετέχοντες τοποθετούνται σε κοινή κλίμακα μονάδων *logit*. Υπάρχει, όμως, μία σημαντική προϋπόθεση στην λογική της *Rasch*: οι ερωτήσεις θα πρέπει να μετρούν μία και μόνο ικανότητα θ , η οποία εμφανίζεται ως κάθετη διακεκομμένη γραμμή στο Σχήμα 1. Αν οι ερωτήσεις μας μετρούσαν διαφορετικές διαστάσεις (διαφορετικές ικανότητες για διαφορετικούς εξεταζόμενους), τότε δύο διαφορετικοί εξεταζόμενοι θα μπορούσαν να πετύχουν το ίδιο σκορ με διαφορετικό τρόπο. Έτσι θα πρέπει αν δούμε την μονοδιαστατικότητα των ερωτημάτων. Πριν από αυ-



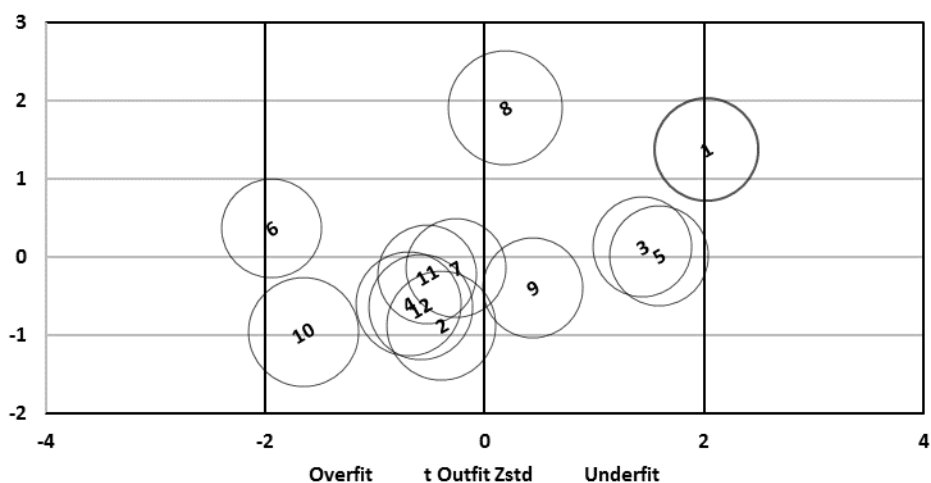
Σχήμα 1. Χρήση της ανάλυσης κατά *Rasch* για την χαρτογράφηση συμμετεχόντων και ερωτημάτων σε κοινή ισοδιαστημική κλίμακα (*Wright's map*).

Η καλή προσαρμογή του μοντέλου *Rasch*

Στον Πίνακα 1 παρουσιάζονται και δύο δείκτες καλής προσαρμογής του μοντέλου Rasch, οι infit και ο outfit, εκφρασμένοι ως οι μέσοι όροι αθροισμάτων τετραγώνων (MNSQ). Ο Outfit είναι ο μέσος όρος των τυπικών διαφορών των παρατηρούμενων τιμών από τις αναμενόμενες τιμές βάσει του Rasch. Ο τύπος του είναι $\frac{1}{N} \sum_{i=1}^N \frac{(U_{ij} - P_{ij})^2}{w_{ij}}$, όπου U_{ij} είναι η παρατηρούμενη επίδοση του μαθητή i στο ερώτημα j , P_{ij} η πιθανότητα ορθής απάντησης του μαθητή i στο ερώτημα j σύμφωνα με το μοντέλο Rasch, και N το μέγεθος του δείγματος. Για κάθε εξεταζόμενο η διαφορά μεταξύ παρατηρούμενης και θεωρητικής τιμής υψώνεται στο τετράγωνο και διαιρείται με τη συνολική διακύμανση του μοντέλου, η οποία είναι $w_{ij} = P_{ij}(1 - P_{ij})$. Ο infit, από την άλλη μεριά, είναι το άθροισμα των τετράγωνων των τυπικών διαφορών μεταξύ των τιμών που παρατηρούνται και αυτών που θα έπρεπε να έχουν παρατηρηθεί με βάση το μοντέλο Rasch (όπως ακριβώς ο outfit), αλλά κάθε ένα από τα τετράγωνα αυτών των τυποποιημένων διαφορών διαιρείται με τη συνολική διακύμανση του μοντέ-

λου. Είναι δηλαδή $\frac{\sum_{i=1}^N (U_{ij} - P_{ij})^2 / w_{ij}}{\sum_{i=1}^N w_{ij}}$.

Ο outfit επηρεάζεται κυρίως από αποκλίσεις σε ερωτήσεις και εξεταζόμενους που είναι στα άκρα της κλίμακας (εύκολες και δύσκολες ερωτήσεις, μαθητές με υψηλή ή χαμηλή ικανότητα). Αντίθετα, ο Infit επηρεάζεται κυρίως από αποκλίσεις στο μεσαίο φάσμα της δυσκολίας των ερωτήσεων ή της ικανότητας των εξεταζομένων. Οι infit και outfit μας βοηθούν να βλέπουμε ποιες ερωτήσεις ή ποιοι εξεταζόμενοι δεν προσαρμόζονται καλά στην λογική της ανάλυσης. Για παράδειγμα, βλέπουμε στον Πίνα-

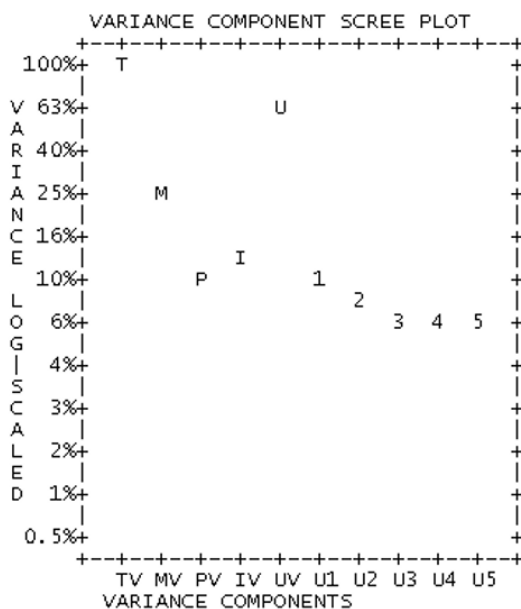


Σχήμα 1. Ο δείκτης outfit για την καλή προσαρμογή των ερωτημάτων στο μοντέλο της ανάλυσης κατά Rasch.

κα 1 ότι η Πρόταση 1 έχει υψηλές τιμές στους δύο αυτούς δείκτες. Αυτό σημαίνει ότι η πρόταση «θα ήθελα να ασχοληθώ επαγγελματικά με τις επιστημονικές κατασκευές» έχει απαντηθεί θετικά και από συμμετέχοντες οι οποίοι σύμφωνα με την λογική της ανάλυσης Rasch θα αναμέναμε να απαντήσουν αρνητικά. Στο Σχήμα 1, βλέπουμε στον οριζόντιο άξονα τις τιμές του outfit και τον κάθετο άξονα την δυσκολία των προτάσεων σε μονάδες logit. Η ερώτηση 1 είναι σχεδόν “underfitted”. Το μέγεθος των κύκλων είναι ανάλογο με τις τυπικές αποκλίσεις του δείκτη που εξετάζουμε.

Η εξασφάλιση της μονοδιαστατικότητας στην Rasch

Ένα σημαντικό χαρακτηριστικό στην μέτρηση κατά Rasch, αλλά όπως λέει και ο Thurstone (1931) σε κάθε μέτρηση στις επιστήμες του ανθρώπου, είναι η εξασφάλιση της μονοδιαστατικότητας, της συνθήκης, δηλαδή, σύμφωνα με την οποία ένα σύνολο ερωτήσεων μετράει την ίδια θεωρητική κατασκευή. Η μονοδιαστατικότητα είναι ιδιαίτερα κρίσιμη στην μέτρηση κατά Rasch, διότι η εν λόγω μέτρηση υποτίθεται ότι είναι «σταθερή» (invariant) και «αντικειμενική» (specific objective). Το πρώτο σημαίνει ότι η δυσκολία κάθε ερώτησης είναι ίδια για όλους τους ερωτώμενους. Το δεύτερο σημαίνει ότι η δυσκολία της κάθε ερώτησης είναι ανεξάρτητη από την ικανότητα (Brentari & Golia, 2007).



Σχήμα 3. Ανάλυση κυρίων συνιστωσών των υπολοίπων της ανάλυσης κατά Rasch.

Στην κλασική θεωρία των μετρήσεων, η απόφαση για τον αριθμό των πιθανών διαστάσεων είναι ενσωματωμένο μέρος της ανάλυσης. Σε ένα μοντέλο ανάλυσης παραγόντων μπορούμε να βρούμε διάφορους τρόπους με διαφορετικό βαθμό πολυπλοκότητας για να αποφασίσουμε τον αριθμό των διαστάσεων (βλ. για παράδειγμα Hair, Anderson, Tatham, & Black, 1995). Στην ανάλυση κατά Rasch, αντίθετα, η προϋπόθεση της μονοδιαστατικότητας και οι μέθοδοι επιβεβαίωσής της είναι ακόμα ζητούμενα. Καταρχάς

οι δείκτες καλής εφαρμογής του μοντέλου Rasch που αναφέραμε σε προηγούμενες

παραγράφους, δηλαδή οι infit και outfit, δεν μπορούν να ανιχνεύσουν την ύπαρξη περισσότερων από μίας διαστάσεων. Αν στην μέτρηση κατά Rasch οι περισσότερες ερωτήσεις μετράνε μία διάσταση, τότε αυτές που μετρούν μια ενδεχόμενη δεύτερη διάσταση παρουσιάζονται απλώς ως ερωτήσεις που δεν προσαρμόζονται στο μοντέλο (όπως η ερώτηση 1 στην δική μας περίπτωση).

Μια τεχνική που μπορούμε να ακολουθήσουμε για τον έλεγχο της μονοδιαστατικότητας είναι να κάνουμε ανάλυση κυρίων συνιστωσών των υπολοίπων (residuals) της Rasch, να χωρίσουμε τις ερωτήσεις σε δύο ομάδες (αυτή των ερωτήσεων τα υπόλοιπα των οποίων σχετίζονται ισχυρά με την πρώτη συνιστώσα και αυτή των ερωτήσεων τα υπόλοιπα των οποίων δεν σχετίζονται ισχυρά με την πρώτη συνιστώσα), να μετρήσουμε την επίδοση (σε logit) των εξεταζόμενων στις δύο ομάδες και να κατασκευάσουμε ένα γράφημα σκεδασμού. Το Σχήμα 2 παρουσιάζει την σχετική ανάλυση κυρίων συνιστωσών για τα δεδομένα μας. Το κάτω μέρος του Πίνακα 2 παρουσιάζει τις πηγές της διακύμανσης στα υπόλοιπα. Παρατηρούμε ότι η διακύμανση που οφείλεται σε μια πρώτη πιθανή συνιστώσα είναι μικρότερη από την διακύμανση που οφείλεται στους συμμετέχοντες ή στα ερωτήματα. Έτσι, θεωρούμε ότι στα δεδομένα μας δεν υπάρχει δεύτερη διάσταση.

Πίνακας 2. Διακύμανση των τυποποιημένων υπολοίπων (standardized residuals) της ανάλυσης κατά Rasch σε μονάδες ιδιοτιμών (eigenvalue units).

Σύμβολο στο Σχήμα 3	Εξήγηση	Ιδιοτιμή	Ποσοστό
[T]otal	Συνολική διακύμανση στις παρατηρήσεις	16,463	100%
[M]easured	Διακύμανση που εξηγείται από τις μετρήσεις	4,4636	27,1%
[P]ersons	Διακύμανση που εξηγείται από τους συμμετέχοντες	2,0053	12,1%
[I]tems	Διακύμανση που εξηγείται από τα ερωτήματα	2,4582	14,9%
[U]nexplained	Διακύμανση που δεν εξηγείται	12,0000	72,9%
[U1]	Ανεξήγητη διακύμανση στην πρώτη συνιστώσα	1,7353	10,5%
[U2]	Ανεξήγητη διακύμανση στην δεύτερη συνιστώσα	1,4622	8,9%
[U3]	Ανεξήγητη διακύμανση στην τρίτη συνιστώσα	1,2943	7,9%
[U4]	Ανεξήγητη διακύμανση στην τέταρτη συνιστώσα	1,2933	7,9%
[U5]	Ανεξήγητη διακύμανση στην πέμπτη συνιστώσα	1,1387	6,9%

Στην σχετική βιβλιογραφία υπάρχουν κι άλλες μέθοδοι ελέγχου της μονοδιαστατικότητας στην Rasch. Ο Verhelst (2001), για παράδειγμα, παρουσιάζει τρεις από αυτές τις μεθόδους: την μέθοδο Martin-Löf (Gustafsson, 1980), το τεστ Van den Wollenberg's Q2-test (Wollenberg, 1982) και την τεχνική "splitter-item" επίσης από τον van den Wollenberg (op cit.). Οι Hambleton και Rovinelli (1986), σε μία και μοναδική εργασία τους, παρουσιάζουν 87 δείκτες ελέγχου της μονοδιαστατικότητας.

Μέσα σε αυτή την πληθώρα των τεστ, κυριαρχούσα άποψη είναι η πρόταση των Reise, Cook και Moore (2015, σελ. 13) η οποία είναι ότι « το κρίσιμο θέμα... δεν είναι το αν τα δεδομένα είναι αρκούντως μονοδιάστατα, αλλά ο βαθμός στον οποίο η [πιθανή] πολυδιαστατικότητα επηρεάζει ή διαστρεβλώνει τον υπολογισμό των παραμέτρων». Ομοίως ο Everett Smith (2002, σελ. 206) γράφει ότι «αντί να ρωτάμε “μονιάστατο ή όχι” [για ένα τεστ], θα έπρεπε να ρωτάμε “σε ποιο βαθμό του συνεχούς [μεταξύ μονοδιαστατικότητας και πολυδιαστατικότητας] η πολυδιαστατικότητα απειλεί την ερμηνεία των ευρημάτων των ερωτήσεων και των εξεταζομένων». Στην επόμενη ενότητα θα προσεγγίσουμε το ζήτημα της μονοδιαστατικότητας από την πλευρά της κλασικής θεωρίας των μετρήσεων.

Η Ανάλυση Κυρίων Συνιστωσών

Μια άλλη μέθοδος αξιολόγησης της μονοδιαστατικότητας των ερωτημάτων που χρησιμοποιήσαμε θα ήταν η Ανάλυση Παραγόντων των αρχικών δεδομένων, μία διαδικασία κατά την οποία οι τιμές των παρατηρήσεων εκφράζονται ως γραμμικά μοντέλα μίας ή περισσότερων διαστάσεων (παραγόντων). Η Ανάλυση Κυρίων Συνιστωσών ανήκει στην οικογένεια της Ανάλυσης Παραγόντων. Στην Ανάλυση Κυρίων Συνιστωσών, η διακύμανση σε ένα σύνολο ερωτημάτων αντικαθίσταται από ένα μικρότερο σύνολο παραγόντων (ορθότερα «συνιστωσών»). Στην παρούσα έρευνα θεωρήσαμε ότι τα «ναι» και τα «όχι» στις 12 ερωτήσεις ήταν στην ουσία εκφράσεις μιας συνεχούς μεταβλητής με τιμές από την απολύτως αρνητική ως την απολύτως θετική γνώμη. Με το σκεπτικό αυτό, χρησιμοποιήσαμε τετραχωρικούς συντελεστές συσχέτισης, οι οποίοι είναι η εναλλακτική λύση για τον συντελεστή r του Pearson στην περίπτωση που οι πιθανές αποκρίσεις είναι μόνο δύο

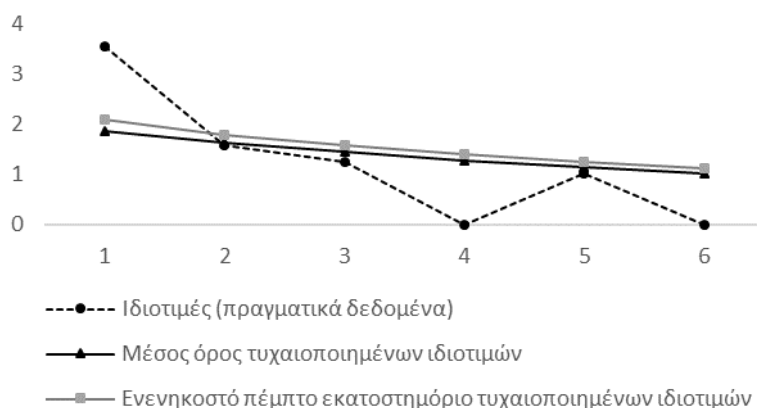
Ο υπολογισμός του τετραχωρικού συντελεστή συνάφειας έχει συζητηθεί εκτενώς στην βιβλιογραφία. (βλ. Divgi, 1979; Mangal, 2002; Bonett & Price, 2005). Η πιο γνωστή μέθοδος υπολογισμού του είναι μέσω του συντελεστή ϕ . Ο ϕ προσεγγίζει τον τετραχωρικό συντελεστή συσχέτισης r_t με το συνημίτονο της ποσότητας $\pi\sqrt{\beta\gamma}/\sqrt{\alpha\delta}\sqrt{\beta\gamma}$, όπου α , β , γ και δ είναι τα ποσοστά των παρατηρήσεων στα τέσσερα κελιά ενός πίνακα δύο γραμμών και δύο στηλών που θεωρούνται ως το πρώτο, το δεύτερο, το τρίτο και το τέταρτο τεταρτοκύκλιο σε έναν υποτιθέμενο κύκλο, ο οποίος χωρίζεται από τις γραμμές διχοτόμησης των κατανομών των δύο συσχετιζόμενων μεταβλητών (βλ. Bouvier, Perry, & Michael, 1954)

Με βάση τους τετραχωρικούς συντελεστές, προχωρήσαμε σε ανάλυση Κυρίων Συνιστωσών (ΑΚΣ), την λογική της οποίας δεν θα αναλύσουμε εδώ λόγω χώρου. Η τιμή του κριτηρίου του Bartlett, το οποίο ελέγχει στην ΑΚΣ την υπόθεση ότι οι μεταβλητές σχετίζονται μεταξύ τους σε βαθμό ικανοποιητικό (συγκεκριμένα ότι ο πίνακας διακυμάνσεων-συνδιακυμάνσεων έχει 1 στην διαγώνιο και 0 εκτός διαγωνίου) ήταν 175,2 και στατιστικώς σημαντική ($p < 0,001$). Αυτό σημαίνει ότι τα δεδομένα μας ήταν κατάλληλα για Ανάλυση Κυρίων Συνιστωσών. Η τιμή του κριτηρίου Kaiser-Meyer-Olkin (KMO), το οποίο αξιολογεί την συνάφεια μεταξύ μεμονωμένων μεταβλητών (partial correlation) σε σχέση προς την συνολική συνάφεια, ήταν ικανοποιητική για

ΑΚΣ (ίση με 0,655).

Στην συνέχεια προχωρήσαμε σε Παράλληλη Ανάλυση (Parallel Analysis),

ώστε να επιλέξουμε έναν αριθμό συνιστωσών που



Σχήμα 3. Parallel Analysis (μέθοδος Horn) για την επιλογή του βέλτιστου αριθμού διαστάσεων.

να εκφράζει καλύτερα τον αρχικό πίνακα διακυμάνσεων-συνδιακυμάνσεων. Ο εισηγητής της Παράλληλης Ανάλυσης, John Horn (1965), πρότεινε την δημιουργία τυχαιοποιημένων συσχετίσεων ανάμεσα στις μεταβλητές και την σύγκριση των ιδιοτιμών των τυχαιοποιημένων συνιστωσών με τις παρατηρούμενες ιδιοτιμές.

Το πρόγραμμα *Factor*, το οποίο χρησιμοποιήσαμε, δημιούργησε 500 τυχαίους πίνακες (Monte Carlo simulations). Στο Σχήμα 3 βλέπουμε πώς μόνο μία από τις παρατηρούμενες συνιστώσες (διακεκομμένη γραμμή) έχει ιδιοτιμή μεγαλύτερη από το 95^ο εκατοστημόριο της κατανομής των τυχαιοποιημένων ιδιοτιμών (γκρι γραμμή). Οι πραγματικές και οι τυχαιοποιημένες ιδιοτιμές παρουσιάζονται λεπτομερώς και στον Πίνακα 3. Έτσι, σύμφωνα με την Παράλληλη Ανάλυση, καταλήξαμε ότι μπορούμε να κρατήσουμε μία και μόνο διάσταση (συνιστώσα) στα δεδομένα μας, η οποία να «εξηγεί» στατιστικά το 29,6% της αρχικής διακύμανσης.

Πίνακας 3. Παρατηρούμενες και τυχαιοποιημένες ιδιοτιμές για την επιλογή του βέλτιστου αριθμού παραγόντων (μέθοδος Horn).

Αριθμός παραγόντων	Ιδιοτιμές (πραγματικά δεδομένα)	Μέσος όρος τυχαιοποιημένων ιδιοτιμών	Ενενηκοστό πέμπτο εκατοστημόριο τυχαιοποιημένων ιδιοτιμών
Παράγοντας 1	3,55701	1,87303	2,08565
Παράγοντας 2	1,58699	1,62990	1,78469
Παράγοντας 3	1,24885	1,44547	1,57284
Παράγοντας 4	1,12.886	1,28782	1,39457
Παράγοντας 5	1,02907	1,15306	1,25317
Παράγοντας 6	0.87807	1,02357	1,11550

Στον Πίνακα 4 που ακολουθεί βλέπουμε τις φορτίσεις της μιας και μοναδικής διάστασης (συνιστώσας) στα 12 ερωτήματα που χρησιμοποιήσαμε. Η επίσκεψη σε πλανητάριο, οι κατασκευές στο σπίτι, η αναζήτηση επιστημονικών γνώσεων στο διαδίκτυο, η γνώση των αστερισμών στον νυχτερινό ουρανό, η επιθυμία των μαθητών να εργαστούν ως επιστήμονες ή μηχανικοί είναι θέματα (ερωτήματα) που σχετίζονται περισσότερο με την μοναδική κοινή διάσταση.

Πίνακας 4. Φορτίσεις της συνιστώσας στα 12 ερωτήματα.

	Προτάσεις	Συντελεστής	Communality
1	Θα ήθελα να ασχοληθώ επαγγελματικά στο μέλλον με τις επιστημονικές κατασκευές	0,182	0,033
2	Αναζητώ επιστημονικές γνώσεις στο διαδίκτυο	0,596	0,356
3	Μπορώ να φτιάξω ένα τηλεσκόπιο με απλά υλικά στο σπίτι	0,294	0,086
4	Μου αρέσει να φτιάχνω πράγματα στο σπίτι που σχετίζονται με την τεχνολογία	0,702	0,493
5	Χρησιμοποιώ τις γνώσεις μου στην αστρονομία για να εξηγήσω διάφορα κοσμικά φαινόμενα	0,307	0,094
6	Μπορώ να αναγνωρίσω τους αστερισμούς στον νυχτερινό ουρανό	0,658	0,433
7	Θα ήθελα να εργαστώ ως επιστήμονας ή μηχανικός	0,570	0,324
8	Έχω διαβάσει βιβλία σε σχέση με το πλανητικό μας σύστημα	0,502	0,252
9	Μπορώ να γράψω τρεις επιπτώσεις της υπερθέρμανσης του πλανήτη	0,507	0,257
10	Έχω επισκεφθεί πλανητάριο	0,787	0,619
11	Στο σπίτι κάνω τεχνολογικές κατασκευές που βλέπω στο διαδίκτυο	0,618	0,382
12	Γνωρίζω τέσσερα διαφορετικά επαγγέλματα που συνδέονται με την NASA	0,478	0,228

Το ζήτημα της αξιοπιστίας

Απαραίτητο στοιχείο σε κάθε μέτρηση είναι και η έννοια της αξιοπιστίας. Αξιοπιστία είναι η πραγματική διακύμανση σε ένα δείγμα συμμετεχόντων διαιρεμένη διά την παρατηρούμενη διακύμανση, η οποία εμπεριέχει και το σφάλμα της μέτρησης. Σε τεστ με διχοτομικές μεταβλητές, η αξιοπιστία μετριέται γενικά με τον συντελεστή των Kuder και Richardson (1937), γενίκευση του οποίου είναι ο δείκτης *alpha* (άλφα) του Cronbach. Ο δείκτης Kuder-Richardson (KR-20), ο οποίος ονομάζεται “Formula 20”,

είναι $r_{kr20} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum pq}{\sigma^2} \right)$, όπου k είναι το πλήθος των ερωτήσεων, p_j είναι το

πλήθος των συμμετεχόντων στο δείγμα που απάντησαν την ερώτηση j σωστά, q_j είναι ο αριθμός των συμμετεχόντων που δεν απάντησαν σωστά την ερώτηση j , ενώ σ^2 είναι η διακύμανση των συνολικών σκορ για όλους τους συμμετέχοντες. Η διάσταση την οποία μετρήσαμε, ο συντελεστής αξιοπιστίας ήταν $r_{kr20}=0,784$. Ένας υψηλός συντελεστής αξιοπιστίας δεν σημαίνει αυτομάτως ότι το τεστ είναι μονοδιάστατος. Αντιστρόφως, αν ένα τεστ είναι μονοδιάστατο, δεν έχει αναγκαστικά υψηλό συντελεστή αξιοπιστίας (Green, Lissitz, and Mulaik, 1977; Cortina, 1993; Sijtsma, 2009). Υπό την προϋπόθεση της μονοδιαστατικότητας, ο *alpha* του Cronbach και ο KR-20 δίνουν ίδια αποτελέσματα.

Συζήτηση

Στην παρούσα εργασία είδαμε πώς η Ανάλυση Κυρίων Συνιστωσών και η ανάλυση κατά Rasch ελέγχουν το ζήτημα της μονοδιαστατικότητας. Οι δύο αυτές ψυχομετρικές τεχνικές ανήκουν σε διαφορετικές λογικές. Η πρώτη ανήκει στην λογική του «πραγματικού σκορ» ενώ η δεύτερη ανήκει στην λογική της θεωρίας των αποκρίσεων (item response theory). Οι δύο τεχνικές μοιάζουν στο ότι στην ουσία είναι δομικές εξισώσεις (structural equation models).

Η Ανάλυση Κυρίων Συνιστωσών (ΑΚΣ) μοιάζει με την γραμμική παλινδρόμηση, με την διαφορά ότι στην ΑΚΣ οι ερευνητές πρέπει να αποφασίσουν τον αριθμό των συνιστωσών και των φορτίσεών τους, θέτοντας κάποιους περιορισμούς. Για παράδειγμα, κάποιες από τις φορτίσεις να είναι μηδέν. Στην ανάλυση κατά Rasch οι δύο λανθάνουσες μεταβλητές είναι η ικανότητα και η δυσκολία. Οι περιορισμοί εδώ είναι ότι οι ερωτήσεις και οι συμμετέχοντες θα πρέπει να έχουν καλή εφαρμογή στη λογική του μοντέλου και, το κυριότερο, η ικανότητα να είναι μονοδιάστατη.

Στην βιβλιογραφία, πολλοί ερευνητές έχουν συζητήσει την σχέση μεταξύ της μη γραμμικής ανάλυσης παραγόντων και της ανάλυσης κατά Rasch (βλ. Holt, Duijn, & Boomsma, 2010; Takane & de Leeuw, 1987; Kamata & Bauer, 2008; Maydeu-Olivares, Cai, & Hernández, 2011; Wirth & Edwards, 2007). Το συμπέρασμα που βγαίνει είναι ότι όταν έχουμε μη γραμμική ανάλυση παραγόντων οι παράμετροι μετατρέπονται σε παραμέτρους της θεωρίας των αποκρίσεων και αντιστρόφως (Kamata & Bauer, 2008).

Οι δύο προσεγγίσεις μοιάζουν και στο ζήτημα της αξιοπιστίας, οι τύποι του υπολογισμού της οποίας είναι ίδιοι στην Ανάλυση Παραγόντων και στην Ανάλυση κατά Rasch. Εκεί που οι δύο προσεγγίσεις διαφέρουν είναι στο ζήτημα ελέγχου της μονοδιαστατικότητας. Στην Ανάλυση Παραγόντων η αναζήτηση κοινών παραγόντων ή συνιστωσών είναι μέρος αυτής καθ' αυτής της ανάλυσης. Στην ανάλυση κατά Rasch η μονοδιαστατικότητα είναι κρίσιμη και οι μέθοδοι για τον έλεγχο της ακόμα συζητούνται.

Γιατί να χρησιμοποιούμε την ανάλυση κατά Rasch και όχι την Ανάλυση Κυρίων Συνιστωσών; Δε θα μπορούσαμε να χρησιμοποιήσουμε την δεύτερη, να εξασφαλίσουμε την μονοδιαστατικότητα και μετά να ιεραρχήσουμε τις ερωτήσεις κατά σειρά δυσκολίας, όπως προτείνουν οι Wu και Adams (2007); Η απάντηση είναι ότι στην ανάλυση κατά Rasch η μέτρηση της δυσκολίας των ερωτημάτων και της ικανότητας των συμμετεχόντων μετριούνται στην ίδια λογαριθμική κλίμακα. Δεύτερον αυτή η κλίμακα είναι ισοδιαστημική και όχι τακτικών τιμών όπως είναι στην Ανάλυση Κυρίων Συνιστωσών. Αυτό εξαλείφει την ανάγκη για κανονικοποίηση των αρχικών τιμών στους παράγοντες, μία διαδικασία που χρησιμοποιείται στην ανάλυση παραγόντων, σύμφωνα με την οποία οι βαθμοί στους παράγοντες κατατάσσονται κατά σειρά μεγέθους και μετά υπόκεινται σε μη γραμμικό μετασχηματισμό για να πάρουν την κανονική κατανομή (βλ. Gibbons, 1983). Τρίτον η μέτρηση κατά Rasch έχει το πλεονέκτημα της ανεξαρτησίας (invariance), όπως καταδεικνύει ο Engelhard (2013). Αυτό σημαίνει ότι στην Rasch η κατανομή της ικανότητας των συμμετεχόντων είναι ανεξάρτητη από την κατανομή της δυσκολίας των ερωτημάτων. Εν κατακλείδι, ο έλεγχος των διαστάσεων ενός τεστ είναι πιο εύκολος με την Ανάλυση Κυρίων Συνιστωσών. Τα μοντέλα Rasch, από την άλλη μεριά (υπό την προϋπόθεση της μονοδιαστατικότητας), είναι «καλύτερα» (φέρουν περισσότερη «πληροφορία»). Η επιλογή μεταξύ των δύο έχει να κάνει με τους σκοπούς της αξιολόγησης και την φύση των δεδομένων.

Βιβλιογραφικές αναφορές

- Beet, E. (2015). *Teaching Astronomy in schools*. Cambridge: Cambridge University Press.
- Bonett, D., & Price, R. (2005). Inferential methods for the tetrachoric correlation coefficient. *Journal of Educational and Behavioral Statistics*, 30(2), 213–225. <http://doi.org/10.3102/10769986030002213>
- Bord, T., & Fox, C. (2015). *Applying the Rasch Model: fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.
- Bouvier, E. A., Perry, N. C., & Michael, W. B. (1954). A study of the error in the cosine-phi approximation to the tetrachoric coefficient of correlation. *Educational and Psychological Measurement*, 14(4), 690–699. <http://doi.org/10.1177/001316445401400413>
- Brentari, E., & Golia, S. (2007). Unidimensionality in the Rasch Model: How to detect and interpret. *Statistica*, 3, 253–261. <http://doi.org/10.6092/issn.1973-2201/3508>
- Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. <http://doi.org/10.1037/0021-9010.78.1.98>
- Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, 44(2), 169–172. <http://doi.org/10.1007/BF02293968>
- Engelhard, G. (2013). *Invariant measurement: using Rasch models in the social, behavioral, and health sciences*. New York, NY: Routledge.
- Gessaroli, M., & de Champlain, A. (2005). Test dimensionality: assessment of. In *Encyclopedia of Statistics in Behavioral Science* (pp. 2014–2020). Wiley and Sons. <http://doi.org/10.1002/0470013192.bsa027>
- Gibbons, A. C. (1983). Primary measures of music audiation scores in an institutionalized elderly population. *Journal of Music Therapy*, 20(1), 21–29.
- Goldstein, H. (1979a). Consequences of using the Rasch model for educational assessment. *British Educational Research Journal*, 5(2), 211–220. Retrieved

from <http://www.jstor.org/stable/1501031>

- Goldstein, H. (1979b). The mystification of assessment. *Forum of the Discussion of New Trends in Education*, 22(1), 14–16. Retrieved from www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/mystification-of-assessment.pdf
- Goldstein, H. (2004). The Education World Cup : international comparisons of student achievement by. *CADMO, Anno XII(2)*, 63–70. Retrieved from <http://eprints.ioe.ac.uk/8786/>
- Goldstein, H. (2015). Rasch measurement: a response to Payanides, Robinson and Tymms, *41(1)*, 176–179. <http://doi.org/10.1002/berj.3170>
- Goldstein, H., & Blinkhorn, S. (1982). The Rasch Model Still Does Not Fit. *British Educational Research Journal*, 8(2), 167–170. <http://doi.org/10.1080/0141192820080207>
- Green, S., Lissitz, R., & Mulaik, S. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37(4), 827–838. <http://doi.org/10.1177/001316447703700403>
- Gustafsson, J. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33, 205–233.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis with readings*. New Jersey: Prentice-Hall Inc.
- Hambleton, R., & Rovinelli, R. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement*, 10(3), 287–302. <http://doi.org/10.1177/014662186030287-16>
- Holt, J. C., Duijn, M. a J. Van, & Boomsma, A. (2010). Scale construction and evaluation in practice: A review of factor analysis versus item response theory applications. *Psychological Test and Assessment Modeling*, 52(3), 272–297. Retrieved from <http://psycnet.apa.org/psycinfo/2010-26531-003>
- Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <http://doi.org/10.1007/BF02289447>
- Iramaneerat, C., Smith, E., & Smith, R. (2008). An introduction to Rasch

- measurement. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 50–71). Thousand Oaks, CA: Sage. <http://doi.org/10.4135/9781412995627>
- Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and Item Response Theory Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 136–153. <http://doi.org/10.1080/10705510701758406>
- Kreiner, S., & Christensen, K. B. (2013). Analyses of model fit and robustness. A new look at the PISA scaling model, underlying ranking of countries according to reading literacy. *Psychometrika*, *79*(2), 210–231. <http://doi.org/10.1007/S11336-013-9347-Z>
- Kuder, G., & Richardson, M. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160. <http://doi.org/10.1007/BF02288391>
- Linacre, J. (1999). Understanding Rasch measurement: estimation methods for Rasch measures. *Journal of Outcome Measurement*, *3*(4), 382–405.
- Linacre, J. (2004). Rasch model estimation: further topics. *Journal of Applied Measurement*, *5*(1), 95–110.
- Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the Fit of Item Response Theory and Factor Analysis Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(3), 333–356. <http://doi.org/10.1080/10705511.2011.581993>
- Panayides, P., Robinson, C., & Tymms, P. (2015). Rasch measurement: a response to Goldstein. *British Educational Research Journal*, *41*(1), 180–182. <http://doi.org/10.1002/berj.3182>
- Pasachoff, J., & Percy, J. (Eds.). (2005). *Teaching and learning astronomy: effective strategies for educators worldwide*. Cambridge: Cambridge University Press.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests [Reprint of the original 1960 work, with a Foreword and Afterword by Benjamin Wright]*. Chicago, IL: University of Chicago Press.
- Reise, S., Cook, K., & Moore, T. (2015). Evaluating the impact of multidimensionality on unidimensional Item Response Theory model parameters. In S. Reise & D. Revicki (Eds.), *Handbook of Item Response Theory*

- Modeling* (pp. 13–40). New York, NY: Routledge.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120.
<http://doi.org/10.1007/s11336-008-9101-0>
- Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and Principal Component Analysis of residuals. *Journal of Applied Measurement*, *3*(2), 205–231.
- Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*(3), 393–408.
<http://doi.org/10.1007/BF02294363>
- Thurstone, L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology*, *26*(3), 249–169. <http://doi.org/10.1037/h0070363>
- Törmäkangas, K. (2011). Advantages of the Rasch measurement model in analysing educational tests: an applicator's reflection. *Educational Research and Evaluation*, *17*(5), 307–320. <http://doi.org/10.1080/13803611.2011.630562>
- Verhelst, N. (2001). Testing the unidimensionality assumption of the Rasch model. *Methods of Psychological Research Online*, *6*(3), 231–271.
- Waugh, R., & Chapman, E. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: what is the difference? Which method is better? *Journal of Applied Measurement*, *6*(1), 80–99.
- Wiberg, M. (2012). Can a multidimensional test be evaluated with unidimensional item response theory? *Educational Research and Evaluation*, *18*(4), 307–320.
<http://doi.org/10.1080/13803611.2012.670416>
- Wollenberg, A. (1982). A simple and effective method to test the dimensionality axiom of the Rasch model. *Applied Psychological Measurement*, *6*(1), 83–91.
- Wright, B. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*(2), 97–116. <http://doi.org/10.2307/1434010>
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions. Retrieved from

[http://media.metrik.de/uploads/incoming/pub/Literatur/von
Winfried/RaschMeasurement_Complete.pdf](http://media.metrik.de/uploads/incoming/pub/Literatur/von%20Winfried/RaschMeasurement_Complete.pdf)