ePublishing
ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL DOCUMENTATION CENTRE

Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση

(2024)

8ο Πανελλήνιο Επιστημονικό Συνέδριο «Ένταξη και Χρήση των ΤΠΕ στην Εκπαιδευτική Διαδικασία»

Ελληνική προσαρμογή και ψυχομετρική επικύρωση της κλίμακας Beginners Computational Thinking Test (BCTt) για την αξιολόγηση της Υπολογιστικής Σκέψης μαθητών/τριών Δημοτικού Σχολείου

Ιωάννης Βουρλέτσης, Παναγιώτης Πολίτης

Βιβλιογραφική αναφορά:

Βουρλέτσης Ι., & Πολίτης Π. (2025). Ελληνική προσαρμογή και ψυχομετρική επικύρωση της κλίμακας Beginners Computational Thinking Test (BCTt) για την αξιολόγηση της Υπολογιστικής Σκέψης μαθητών/τριών Δημοτικού Σχολείου. *Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση*, 619–632. ανακτήθηκε από https://eproceedings.epublishing.ekt.gr/index.php/cetpe/article/view/8477

https://epublishing.ekt.gr | **e-Εκδότης:** ΕΚΤ | **Πρόσβαση:** 22/01/2026 06:05:26

Ελληνική προσαρμογή και ψυχομετρική επικύρωση της κλίμακας Beginners Computational Thinking Test (BCTt) για την αξιολόγηση της Υπολογιστικής Σκέψης μαθητών/τριών Δημοτικού Σχολείου

Ιωάννης Βουρλέτσης, Παναγιώτης Πολίτης
vourlets@uth.gr, ppol@uth.gr
Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Πανεπιστήμιο Θεσσαλίας

Περίληψη

Η Υπολογιστική Σκέψη (ΥΣ) αφορά διαδικασίες της σκέψης που είναι εφαρμόσιμες σε πλήθος πεδίων και θεωρείται πολύτιμο σύνολο δεξιοτήτων για τους/τις μαθητές/τριες του 21ου αιώνα. Ωστόσο, περιορισμένη παραμένει η έρευνα που εστιάζει στα εργαλεία αξιολόγησής της. Το Beginners Computational Thinking Test (BCTt) είναι ένα έγκυρο και αξιόπιστο εργαλείο για την αξιολόγηση εννοιών της ΥΣ μαθητών/τριών ηλικίας 5 ως 10 ετών. Η παρούσα μελέτη, μέσα από τη ελληνική προσαρμογή και την ψυχομετρική επικύρωση της κλίμακας BCTt, διερευνά τον βαθμό ανάπτυξης εννοιών της ΥΣ μαθητών/τριών Α' ως Γ' δημοτικού σχολείου. Η εξέταση των ψυχομετρικών ιδιοτήτων της προσαρμοσμένης κλίμακας με δείγμα 517 μαθητών/τριών Α' ως Γ' τάξης έδειξε ότι αποτελεί έγκυρο και αξιόπιστο εργαλείο αξιολόγησης εννοιών της ΥΣ κυρίως για την Α' και Β' τάξη. Τα ευρήματά μας υπογραμμίζουν την αναπτυξιακή προοπτική στην καλλιέργεια των εννοιών της ΥΣ και συμβάλλουν στην ισότιμη συμμετοχή και εξέλιξη στην ψηφιακή εποχή.

Λέξεις κλειδιά: Υπολογιστική Σκέψη (ΥΣ), Αξιολόγηση, Πολιτισμική προσαρμογή, Ψυχομετρική επικύρωση, Δημοτικό σχολείο

Εισαγωγή

Ο όρος *Υπολογιστική Σκέψη* (Computational Thinking), στο εξής ΥΣ, εμφανίστηκε για πρώτη φορά το 1980, όταν ο Seymour Papert την περιέγραψε ως μια διανοητική δεξιότητα που αποκτούν τα παιδιά μέσω του προγραμματισμού (Papert, 1980). Η ιδέα της ΥΣ ήταν ακόμη παλαιότερη, αλλά προκάλεσε εκτεταμένη ακαδημαϊκή συζήτηση το 2006, όταν η Jeannette Wing την περιέγραψε ως ένα σύνολο νοητικών εργαλείων για την επίλυση προβλημάτων και ως εξίσου σημαντική με την ανάγνωση, τη γραφή και την αριθμητική (Wing, 2006). Η ΥΣ σταδιακά ενσωματώθηκε στην υποχρεωτική εκπαίδευση αρκετών χωρών και ο προγραμματισμός του υπολογιστή κυριάρχησε ανάμεσα στις στρατηγικές καλλιέργειας (Flórez et al., 2017) και αξιολόγησής της (Adams et al., 2019).

Η εστίαση της έρευνας, ωστόσο, που αφορά στα εργαλεία αξιολόγησης της ΥΣ συχνά αποτελεί τροχοπέδη για την ένταξή της στην υποχρεωτική εκπαίδευση (Babazadeh & Negrini, 2022), καθώς ελάχιστα από τα εργαλεία αξιολόγησης της ΥΣ που προτείνονται σε αυτή είναι ανεξάρτητα από ένα περιβάλλον προγραμματισμού και κατάλληλα για χρήση με μικρότερους μαθητές του δημοτικού. Το *Beginners Computational Thinking test* (BCTt· Zapata-Cáceres et al., 2020, 2021) αίρει αυτούς τους περιορισμούς, καθώς δεν απαιτεί τη χρήση προγραμματιστικού περιβάλλοντος, είναι σχεδιασμένο για μαθητές ηλικίας 5 ως 10 ετών και μπορεί να χρησιμοποιηθεί τόσο ως εργαλείο προέλεγχου (pre-test) όσο και μετέλεγχου (post-test).

Θεωρητικό πλαίσιο

Η Υπολογιστική Σκέψη

Η ΥΣ φαίνεται ότι έχει τις ρίζες της στη μέθοδο επίλυσης προβλημάτων τεσσάρων σταδίων του Polya (1945) που μπορεί να εφαρμοστεί στα μαθηματικά και σε άλλες επιστήμες. Παρόλα αυτά, τις μελλοντικές εξελίξεις καθόρισε το 2006 η Wing με τη δήλωσή της ότι «η ΥΣ είναι κάτι περισσότερο από το να είσαι σε θέση να προγραμματίζεις έναν υπολογιστή, καθώς απαιτεί σκέψη σε πολλαπλά επίπεδα αφαίρεσης» (Wing, 2006, σ. 34). Σταδιακά επιτεύχθηκε σχετική συναίνεση για το περιεχόμενο του όρου της ΥΣ, που περιλαμβάνει τη διατύπωση προβλημάτων και την ανάπτυξη λύσεών τους με τρόπο που μπορούν να εκτελεστούν με ελάχιστα προαπαιτούμενα από έναν άνθρωπο ή μια μηχανή (Grover & Pea, 2018; Wing, 2011). Στο πεδίο της εκπαίδευσης, το τρισδιάστατο μοντέλο των Brennan και Resnick (2012) ήταν ιδιαίτερα επιδραστικό στον ορισμό της ΥΣ και σύμφωνα μ' αυτό, η ΥΣ αποτελείται από έννοιες, πρακτικές και στάσεις/θεωρήσεις.

Οι Annamalai et al. (2022) στην πρόσφατη βιβλιογραφική τους επισκόπηση συμπέραναν ότι οι κύριες διαστάσεις της ΥΣ περιλαμβάνουν την αφαίρεση, την αποσύνθεση του προβλήματος, την εκσφαλμάτωση και αξιολόγηση, τους αλγόριθμους και τη γενίκευση. Ακόμα, η έρευνά τους έδειξε ότι η ΥΣ μπορεί να ενισχύει τις γενικές δεξιότητες επίλυσης προβλημάτων. Τέλος, είναι σημαντικό να σημειωθεί ότι η ΥΣ, αν και συχνά συνδέεται με την τεχνολογία, δεν την απαιτεί πάντοτε (Hazzan et al., 2020).

Αξιολόγηση της Υπολογιστικής Σκέψης

Παρά το συνεχώς αυξανόμενο πλήθος των μεθόδων αξιολόγησης της ΥΣ, φαίνεται πως περιορισμένη είναι η έρευνα σχετικά με την εγκυρότητα και την αποτελεσματική τους ενσωμάτωση στα εκπαιδευτικά πλαίσια (Román-González et al., 2019). Ακόμη, διαπιστώνεται ότι υπάρχει ανάγκη για μεγάλης κλίμακας επικύρωση και εφαρμογή τους (Tang et al., 2020; Lu et al., 2022) και κυρίως για μικρότερης ηλικίας μαθητές/τριες (Cutumisu et al., 2019; Poulakis & Politis, 2021).

Μεταξύ των υφιστάμενων εργαλείων αξιολόγησης της ΥΣ, πολύ λίγα είναι ανεξάρτητα από ένα περιβάλλον προγραμματισμού (Zapata-Cáceres et al., 2020) και κατάλληλα για χρήση ως εργαλεία προέλεγχου και μετέλεγχου. Ακόμη και ευρέως χρησιμοποιούμενες προσεγγίσεις την αξιολόγηση της ΥΣ, όπως η δοκιμασία Bebras (bebras.org), που απευθύνονται σε ευρύ φάσμα ηλικιών μαθητών/τριών, συχνά απαιτούν τέτοιο βαθμό αναγνωστικής κατανόησης, που θα μπορούσε να επηρεάσει την αξιολόγηση δεξιοτήτων της ΥΣ, ειδικά στα πρώιμα στάδια ανάπτυξης (Zapata-Cáceres et al., 2024). Το Computational Thinking Test (CTt· Román-González, 2015; Román-González et al., 2017) αποτελεί ένα αξιόπιστο και έγκυρο εργαλείο αξιολόγησης της ΥΣ μαθητών/τριών ηλικίας 10 ως 16 ετών και αποτέλεσε τη βάση για τον σχεδιασμό του BCTt.

Το πρωτότυπο Beginners Computational Thinking Test

Το BCTt βασίζεται στο CTt, αλλά έχει προσαρμοστεί για μικρότερους/ες μαθητές/τριες. Όπως και το CTt, το BCTt μπορεί να αξιοποιείται ως εργαλείο προέλεγχου και μετέλεγχου, είναι ανεξάρτητο από κάποιο περιβάλλον προγραμματισμού και «μπορεί να χρησιμοποιηθεί σε μαθητές του δημοτικού σχολείου, ιδιαίτερα στις πρώτες τάξεις (5 ως 10 ετών)» (Zapata-Cáceres et al., 2020, σ. 1913).

Μετά την ψυχομετρική του επικύρωση, το BCTt αποτελείται από 25 ερωτήσεις, χωρισμένες σε έξι σύνολα, καθένα από τα οποία καλύπτει μια ξεχωριστή υπολογιστική έννοια: τις *ακολουθίες*, τους *απλούς* και *φωλιασμένους* βρόχους και τις συνθήκες *αν-τότε*, *αν-τότε-αλλιώς* και

όσο. Για κάθε ερώτηση, μόνο μία από τις τέσσερις δυνατές απαντήσεις είναι σωστή. Η ψυχομετρική επικύρωση του BCTt έδειξε ότι πρόκειται για ένα αξιόπιστο και έγκυρο όσον αφορά τον σχεδιασμό και το περιεχόμενό του εργαλείο, ισορροπημένης δυσκολίας αλλά και ότι η δυσκολία του αυξάνεται προοδευτικά. Τα αρχικά ερωτήματα φαίνεται πως είναι εύκολα για τους μεγαλύτερους/ες μαθητές/τριες, καθιστώντας το πιο κατάλληλο για μαθητές/τριες των μικρότερων τάξεων του δημοτικού σχολείου. Η αξιοπιστία του εργαλείου ήταν επίσης υψηλότερη για τις μικρότερες τάξεις.

Διαφορές ηλικίας και τάξης

Η ανάγκη για τον σχεδιασμό προσαρμοσμένων σε ηλικιακές ομάδες προσεγγίσεων οδήγησε στη διερεύνηση διαφορών στον βαθμό ανάπτυξης συνιστωσών της ΥΣ μεταξύ μαθητών/τριών διαφορετικών ηλικιών με αξιοποίηση διαφορετικών εργαλείων και προσεγγίσεων. Άλλωστε, έχει φανεί πως οι δεξιότητες κωδικοποίησης και προγραμματισμού είναι πιθανό να συνδέονται με την γνωστική ωριμότητα (cognitive maturity· Kim et al., 2021) και να χρειάζονται αρκετό χρόνο για την ολοκληρωμένη τους ανάπτυξη (Atmatzidou & Dimitriadis, 2016).

Οι Jiang και Wong (2021) εξέτασαν τόσο τις αναπτυξιακές διαφορές όσο και τις διαφορές φύλου στις δεξιότητες ΥΣ 197 μαθητών ηλικίας 9 ως 13 ετών. Τα ευρήματά τους έδειξαν ότι οι μεγαλύτεροι μαθητές/τριες επετύγχαναν σταθερά υψηλότερες επιδόσεις από τους/τις νεότερους/ες συμμαθητές/τριές τους. Οι Polat et al. (2021) υλοποίησαν ένα πρόγραμμα ΥΣ 9 εβδομάδων για 1 ώρα την εβδομάδα με 328 μαθητές της Ε΄ και ΣΤ΄ τάξης (ηλικίας 10–12 ετών). Χρησιμοποιώντας το περιβάλλον οπτικού προγραμματισμού Scratch, διαπίστωσαν ότι οι μαθητές της ΣΤ΄ τάξης είχαν καλύτερες επιδόσεις από τους μαθητές της Ε΄ τάξης, αποδίδοντας τη διαφορά στο γεγονός ότι οι πρώτοι συμμετείχαν σε περιπλοκότερα έργα.

Στο πλαίσιο της μεγάλης κλίμακας ανάλυσης των Román-González et al. (2017), το CTt χρησιμοποιήθηκε για την αξιολόγηση των δεξιοτήτων ΥΣ 1.251 Ισπανών/Ισπανίδων μαθητών/τριών από την Ε΄ τάξη του δημοτικού σχολείου ως την Α΄ Λυκείου. Οι ερευνητές συμπέραναν ότι η ΥΣ, ως δεξιότητα επίλυσης προβλημάτων, συνδέεται άμεσα με τη γνωστική ωριμότητα των μαθητών. Τέλος, οι Rijke et al. (2018) διεξήγαγαν έρευνα με 200 μαθητές/τριες δημοτικού σχολείου ηλικίας 6 ως 12 ετών, εφαρμόζοντας διδακτική παρέμβαση με δραστηριότητες χωρίς χρήση υπολογιστών. Τα αποτελέσματά τους έδειξαν ότι η ηλικία των συμμετεχόντων μπορεί να επηρεάσει την ανάπτυξη συγκεκριμένων δεξιοτήτων ΥΣ.

Διαφορές φύλου

Οι μελέτες που εξετάζουν την επίδραση του φύλου στην ανάπτυξη των δεξιοτήτων ΥΣ στους/στις μαθητές/τριες έχουν καταλήξει σε αντικρουόμενα αποτελέσματα. Ειδικότερα, οι Polat et al. (2021) μελέτησαν 328 μαθητές ηλικίας 10-12 ετών και κατέληξαν στο συμπέρασμα ότι τα αγόρια είχαν καλύτερες επιδόσεις από τα κορίτσια στις δεξιότητες ΥΣ. Ωστόσο, το μέγεθος της επίδρασης του φύλου ήταν μικρό, υποδηλώνοντας ότι το χάσμα μεταξύ των φύλων μπορεί να αμβλύνεται. Οι Román-González et al. (2017) επίσης εντόπισαν στατιστικά σημαντικές διαφορές στις βαθμολογίες μεταξύ αγοριών και κοριτσιών από την Ε΄ δημοτικού ως την Α΄ Λυκείου με μικρό μέγεθος επίδρασης του φύλου. Οι μέσες βαθμολογίες ήταν υψηλότερες στην ομάδα των αγοριών μόνο καθώς προχωρούσαν πέρα από την Α΄ Γυμνασίου. Οι Mouza et al. (2020), αφού ανέλυσαν τόσο ποσοτικά όσο και ποιοτικά δεδομένα από 238 μαθητές/τριες της Δ΄ ως ΣΤ΄ τάξης σε διάστημα 2,5 ετών, παρατήρησαν διαφορές στα αποτελέσματα ΥΣ μεταξύ των φύλων, αποδίδοντας τη διαφορά στο «χάσμα πείρας» (Mouza et al., 2020, σ. 1050), δηλαδή στη μεγαλύτερη πείρα των αγοριών με τους υπολογιστές.

Επιπλέον, διαφορές που σχετίζονται με το φύλο έχει εντοπιστεί από έρευνα με μικρότερους/ες μαθητές/τριες και προγράμματα ρομποτικής. Οι Sullivan και Bers (2013) εργάστηκαν με 53 μαθητές/τριες νηπιαγωγείου σε ένα πρόγραμμα ρομποτικής έξι συνεδριών και κατέγραψαν υψηλότερες βαθμολογίες των αγοριών μόνο στην υπολογιστική έννοια των συνθηκών. Οι ίδιοι ερευνητές, σε πιλοτική τους έρευνα με 45 μαθητές/τριες ηλικίας 4 ως 7 ετών εντόπισαν στατιστικά σημαντικές διαφορές μεταξύ αγοριών και κοριτσιών μόνο σε περισσότερο απαιτητικά έργα, που περιελάμβαναν βρόχους και αισθητήρες (Sullivan & Bers, 2016). Οι Angeli και Valanides (2020) συμπέραναν επίσης ότι το φύλο των 50 μαθητών/τριών νηπιαγωγείου ηλικίας 5 ως 6 ετών που εργάζονταν με το Bee-Bot δεν επηρέασε τις δεξιότητες προγραμματισμού τους, αλλά φαίνεται πως μπορεί να υπάρχουν επιδράσεις του φύλου στις διαστάσεις ΥΣ των ακολουθιών και της αποσύνθεσης (Angeli & Georgiou, 2023).

Αναφορικά, τέλος, με την αλληλεπίδραση μεταξύ ηλικίας και φύλου, οι Jiang και Wong (2021) συμπέραναν ότι σε όλη τη διάρκεια της καλλιέργειας δεξιοτήτων ΥΣ, το φύλο 197 μαθητών/τριών ηλικίας 9 ως 13 ετών δεν επέδρασε σημαντικά στην καλλιέργεια των δεξιοτήτων. Ωστόσο, οι Rijke et al. (2018) διαπίστωσαν ότι μετά από την ηλικία των 9,5 ετών, οι μαθήτριες υπερτερούσαν των μαθητών σε έργα που απαιτούσαν νοητικές αφαιρέσεις.

Μεθοδολογία

Ερευνητικά ερωτήματα

Βασικός στόχος της έρευνάς μας υπήρξε η διερεύνηση του βαθμού ανάπτυξης των εννοιών της ΥΣ μαθητών/τριών Α' ως Δ' δημοτικού μέσω της προσαρμοσμένης χρήσης της κλίμακας BCTt. Έμφαση δόθηκε στην εξέταση της εγκυρότητας και της αξιοπιστίας της προσαρμοσμένης κλίμακας, ενώ εξετάστηκαν και τα χαρακτηριστικά των ερωτημάτων της. Ειδικότερα, τέθηκαν τα εξής ερευνητικά ερωτήματα:

1. Είναι η ελληνική έκδοση του BCTt έγκυρη κλίμακα για την αξιολόγηση εννοιών της ΥΣ μαθητών/τριών Α' ως Δ΄τάξης δημοτικού σχολείου;
2. Είναι η ελληνική έκδοση του BCTt αξιόπιστη κλίμακα για την αξιολόγηση εννοιών της ΥΣ μαθητών/τριών Α' ως Δ΄τάξης δημοτικού σχολείου;

Καίριας σημασίας είναι ακόμη η ψυχομετρική επικύρωση εργαλείων αξιολόγησης της ΥΣ που λαμβάνουν υπόψη το φύλο των συμμετεχόντων, ενώ φαίνεται πως απουσιάζουν προς το παρόν από τη διεθνή βιβλιογραφία μελέτες διερεύνησης της επίδρασης της τάξης/ηλικίας και του φύλου στην ανάπτυξη των δεξιοτήτων της ΥΣ (Angeli & Giannakos, 2020). Με βάση τα παραπάνω, τέθηκαν επιπλέον οι εξής στόχοι:

3. Υπάρχουν διαφορές στον βαθμό ανάπτυξης εννοιών της ΥΣ μεταξύ μαθητών/τριών διαφορετικής τάξης φοίτησης;
4. Υπάρχουν διαφορές στον βαθμό ανάπτυξης εννοιών της ΥΣ μεταξύ μαθητών και μαθητριών;
5. Υπάρχει αλληλεπίδραση της τάξης φοίτησης και του φύλου στον βαθμό ανάπτυξης εννοιών της ΥΣ;

Συμμετέχοντες

Προκειμένου να εξασφαλίσουμε αντιπροσωπευτικό δείγμα μαθητών/τριών Α' ως Δ' τάξης δημοτικού σχολείου, πραγματοποιήσαμε δειγματοληψία δύο σταδίων. Κατά το πρώτο στάδιο, διενεργήσαμε τυχαία δειγματοληψία, κατά την οποία οι δειγματοληπτικές ομάδες είχαν πιθανότητες ανάλογες του μεγέθους τους (probability proportional to size· PPS), όπως

προτείνεται στην περίπτωση σημαντικών διαφοροποιήσεων στα μεγέθη των δειγματοληπτικών μονάδων (Cheung, 2014). Ειδικότερα, μετά από ταξινόμηση των 13 περιφερειακών διευθύνσεων πρωτοβάθμιας και δευτεροβάθμιας εκπαίδευσης με κριτήριο το πλήθος των εποπτευόμενων σχολικών μονάδων, επιλέχθηκε εκείνη της Αττικής. Κατά το δεύτερο στάδιο, εφαρμόζοντας απλή τυχαία δειγματοληψία (simple random sampling· SRS), στην οποία οι δειγματοληπτικές μονάδες επιλέγονται από έναν πληθυσμό τυχαία, έχοντας ίσες πιθανότητες (Singh, 2003), επιλέξαμε τυχαία 5 δημοτικά σχολεία. Η ολοκλήρωση των δύο σταδίων της δειγματοληπτικής διαδικασίας οδήγησε σε δείγμα 673 μαθητών και μαθητριών. Μετά από τη διαπίστωση φαινομένου οροφής στις βαθμολογίες των μαθητών/τριών της Δ΄ τάξης, όπως αναλύεται στο κεφάλαιο «Ανάπτυξη Εννοιών ΥΣ μαθητών/τριών», προχωρήσαμε στον αποκλεισμό των 156 μαθητών/τριών της Δ΄ τάξης και τελικά, το δείγμα μας αποτέλεσαν 517 μαθητές και μαθήτριες των Α΄, Β΄ και Γ΄ τάξεων (βλ. Πίνακα 1).

Πίνακας 1. Κατανομή των συμμετεχόντων ανά τάξη και φύλο

Τάξη	Μαθητές (*n*)	Μαθήτριες (*n*)	Σύνολο	Μαθητές (%)	Μαθήτριες (%)
Α΄	63	97	160	39,4	60,6
Β΄	77	95	172	44,8	55,2
Γ΄	88	97	185	47,6	52,4
Σύνολο	228	289	517	44,1	55,9

Διαδικασία συλλογής και ανάλυσης των δεδομένων

Μετά την εξασφάλιση των απαραίτητων αδειών, πραγματοποιήθηκε η συλλογή των δεδομένων από τον πρώτο συγγραφέα κατά τον Μάιο και τον Ιούνιο του 2022. Οι μαθητές/τριες χρειάστηκαν 35 (Γ' τάξη) ως 55 λεπτά (Α' και Β' τάξη) για να ολοκληρώσουν τη δοκιμασία. Κατά την κωδικοποίηση των απαντήσεων, σε καθεμία από τις 25 ερωτήσεις αποδόθηκε 1 βαθμός για καθεμία σωστή απάντηση και δεν αποδόθηκε βαθμός για καθεμία λανθασμένη. Η μέγιστη, επομένως, βαθμολογία ήταν 25 βαθμοί και η ελάχιστη 0 βαθμοί.

Για τη μετάφραση της πρωτότυπης κλίμακας BCTt στην ελληνική γλώσσα και την πολιτισμική της προσαρμογή, ακολουθήσαμε την προτεινόμενη από τη διεθνή βιβλιογραφία μεθοδολογία (Arafat et al. 2016; Beaton et al., 2000; Boateng et al., 2018; Borsa et al., 2012) με μετάφραση από την αγγλική γλώσσα στην ελληνική (forward translation) και στη συνέχεια από την ελληνική στην αγγλική (back translation).

Κατά τη διερεύνηση του βαθμού εγκυρότητας της προσαρμοσμένης κλίμακας, αρχικά εξετάστηκε η εγκυρότητα του περιεχομένου (content validity) από έξι ειδικούς του πεδίου των Τεχνολογιών της Πληροφορίας και των Επικοινωνιών (ΤΠΕ) στην εκπαίδευση. Επιπλέον, εξετάστηκε η εγκυρότητα εννοιολογικής κατασκευής (construct validity) μέσα από τη διενέργεια επιβεβαιωτικής ανάλυσης παραγόντων (confirmatory factor analysis· Hurley et al., 1997) με εφαρμογή της μεθόδου εκτίμησης των βεβαρυμμένων ελαχίστων τετραγώνων με διόρθωση ως προς τον μέσο όρο και τη διακύμανση (weighted least square mean and variance adjusted· WLSMV· Passos et al., 2023). Για την αξιολόγηση του βαθμού προσαρμογής της κλίμακας, υπολογίστηκε η τιμή της αναλογίας χ^2/df (Kline, 2011; Wheaton et al., 1977) και των δεικτών CFI, TLI και NFI (Bentler & Bonett, 1980; Byrne, 1994). Τέλος, υπολογίστηκαν οι τιμές του μέτρου Kaiser-Meyer-Olkin (KMO· Kaiser, 1974) και του δείκτη σφαιρικότητας του Bartlett (Bartlett's test of sphericity· Bartlett, 1951), με σκοπό την αξιολόγηση του βαθμού επάρκειας της δειγματοληψίας και τη δυνατότητα να προκύψουν διακριτοί και αξιόπιστοι παράγοντες.

Για τη διερεύνηση των χαρακτηριστικών των ερωτημάτων της προσαρμοσμένης κλίμακας, αρχικά αξιολογήσαμε το επίπεδο δυσκολίας κάθε ερώτησής της, υπολογίζοντας τον δείκτη δυσκολίας των ερωτήσεων (item difficulty index· IDI· Barnard, 1999) και έπειτα ολόκληρης της κλίμακας. Προκειμένου να διερευνήσουμε την ικανότητα των ερωτήσεων να διακρίνουν μαθητές/τριες διαφορετικών επιπέδων επίδοσης, υπολογίσαμε τον δείκτη διακριτικής ικανότητας (item discrimination index· Azevedo et al., 2019; Hingorjo & Jaleel, 2012). Η αξιολόγηση της εσωτερικής συνέπειας της προσαρμοσμένης κλίμακας BCTt και των υποκλιμάκων της πραγματοποιήθηκε με υπολογισμό του δείκτη Kuder-Richardson (KR20· Kuder & Richardson, 1937) και της μέσης τιμής ενδοστοιχειακών συσχετίσεων των υποκλιμάκων (mean inter-item correlations· MIC), δηλαδή των συσχετίσεων ανάμεσα σε όλα τα δυνατά ζευγάρια των στοιχείων κάθε υποκλίμακας, όπως προτείνεται για μικρής έκτασης κλίμακες (Briggs & Cheek, 1986). Για την αξιολόγηση της αξιοπιστίας εξέτασης-επανεξέτασης της προσαρμοσμένης κλίμακας, υπολογίστηκε ο συντελεστής ενδοταξικής συσχέτισης (Intraclass Correlation Coefficient· ICC· (Streiner et al., 2015).

Προκειμένου να διερευνήσουμε τον βαθμό ανάπτυξης των εννοιών της ΥΣ των μαθητών/τριών του δείγματος, υπολογίσαμε περιγραφικά στατιστικά δεδομένα (μέση τιμή, τυπική απόκλιση) των βαθμολογιών τους. Ακόμα, για τη διερεύνηση ενδεχόμενων διαφορών στις βαθμολογίες της δοκιμασίας μεταξύ μαθητών/τριών διαφορετικών τάξεων πραγματοποιήσαμε ανάλυση διακύμανσης κατά έναν παράγοντα (one-way ANOVA) με τη μέθοδο του Welch (Rasch et al., 2011) και πολλαπλές συγκρίσεις των μέσων τιμών ανά ζεύγη με τη μέθοδο Games-Howell (Toothaker, 1993), ενώ για τη διερεύνηση ενδεχόμενων διαφορών ως προς το φύλο, διενεργήσαμε τον στατιστικό έλεγχο *t*-test ανεξάρτητων δειγμάτων. Ακόμα, υπολογίσαμε την τιμή του συντελεστή η^2 και *d* του Cohen (Cohen, 1977) για τη μέτρηση του βαθμού επίδρασης της τάξης φοίτησης και του φύλου στον βαθμό ανάπτυξης των εννοιών της ΥΣ. Σύμφωνα με τον Cohen (1977), τιμές του συντελεστή $\eta^2 < 0{,}06$ υποδεικνύουν μικρό μέγεθος επίδρασης, ενώ τιμές $\eta^2 \geq 0{,}14$ υποδεικνύουν μεγάλο μέγεθος επίδρασης. Αναφορικά με τις τιμές του συντελεστή d, όταν $d \approx 0.20$ θεωρούμε ότι το μέγεθος επίδρασης είναι μικρό, όταν $d \approx 0.50$ ότι το μέγεθος είναι μέτριο και όταν $d \approx 0.80$ ότι είναι μεγάλο (Cohen, 1977). Τέλος, η διερεύνηση της αλληλεπίδρασης της τάξης και του φύλου στην κατάκτηση εννοιών της ΥΣ έγινε με εφαρμογή ανάλυση διακύμανσης κατά δύο παράγοντες (two-way ANOVA· Kim, 2014) με πολλαπλές συγκρίσεις και υπολογίστηκε η τιμή συντελεστή μεγέθους επίδρασης η_p^2 (Cohen, 1977), για τη διερεύνηση του βαθμού επίδρασης της αλληλεπίδρασης των παραγόντων στην ερμηνεία της διακύμανσης στις βαθμολογίες της δοκιμασίας μεταξύ των υποκειμένων της έρευνας διαφορετικών τάξεων και φύλου. Οι τιμές του συντελεστή η_p^2 ερμηνεύονται με τον ίδιο τρόπο, όπως εκείνες του συντελεστή η^2.

Αποτελέσματα

Μετάφραση και πολιτισμική προσαρμογή

Κατά τη μετάφραση από την αγγλική γλώσσα στην ελληνική, δύο ανεξάρτητοι, δίγλωσσοι μεταφραστές μετέφρασαν το κείμενο στη μητρική τους γλώσσα (ελληνική). Στη συνέχεια, επιλύσαμε μικρές αποκλίσεις μεταξύ των δύο εκδοχών και με τη συμβολή ενός ακόμη μεταφραστή οδηγηθήκαμε στην πρώτη μεταφρασμένη στα ελληνικά εκδοχή του BCTt. Έπειτα, η έκδοση αυτή μεταφράστηκε στην αγγλική γλώσσα από δύο ανεξάρτητους μεταφραστές, που έχουν ως μητρική γλώσσα την αγγλική. Η διαδικασία αυτή λειτουργεί ως «πρόσθετος ποιοτικός έλεγχος» (Borsa et al., 2012, σ. 426) και στοχεύει στην αποκάλυψη παρανοήσεων και ασαφών διατυπώσεων (Beaton et al., 2000; Borsa et al., 2012; Tsang et al., 2017).

Τέλος, αφού φροντίσαμε για την προσαρμογή του λεξιλογίου της κλίμακας σε μορφή οικεία για μαθητές/τριες των μικρών τάξεων του δημοτικού σχολείου, δημιουργήσαμε μια προ-τελική έκδοση της ελληνικής έκδοσης της κλίμακας. Αυτή χορηγήθηκε πιλοτικά σε 43 μαθητές/τριες Α΄ και Γ΄ τάξης, διαφορετικούς/ές από εκείνους/ες του δείγματός μας. Δεν κρίθηκε απαραίτητο να γίνουν περαιτέρω τροποποιήσεις και η ίδια έκδοση αποτέλεσε την τελική έκδοση της προσαρμοσμένης κλίμακας.

Εγκυρότητα

Εγκυρότητα περιεχομένου

Όλα τα στοιχεία της προσαρμοσμένης κλίμακας έλαβαν τιμή δείκτη εγκυρότητας περιεχομένου ερώτησης (I-CVI) μεγαλύτερο του 0,78, όπως προτείνεται από τον Lynn (1986). Τα αποτελέσματα έδειξαν επίσης ομοφωνία μεταξύ των ειδικών του πεδίου των ΤΠΕ στην εκπαίδευση σχετικά με τη συνολική συνάφεια των στοιχείων της κλίμακας, καθώς η τιμή του δείκτη εγκυρότητας σε επίπεδο κλίμακας (S-CVI/Ave= 0,99 και S-CVI/UA= 0,92) πληροί τις προτεινόμενες τιμές της διεθνούς βιβλιογραφίας (Lynn, 1986; Shi et al., 2012).

Εγκυρότητα εννοιολογικής κατασκευής

Τα αποτελέσματα της επιβεβαιωτικής ανάλυσης παραγόντων για τους 6 παράγοντες (υποκλίμακες της προσαρμοσμένης κλίμακας) αποτελούμενους από 25 διχοτομικά στοιχεία (σωστή και λανθασμένη απάντηση) για τις τάξεις Α΄, Β΄ και Γ΄ συνοψίζονται στον Πίνακα 2. Εξετάζοντας τις τιμές των δεικτών προσαρμογής, συμπεραίνουμε ότι το μοντέλο παρουσιάζει καλή προσαρμογή για την Α΄ τάξη. Ειδικότερα, η αναλογία χ^2/df είναι μικρότερη της τιμής 3, γεγονός που υποδηλώνει καλή προσαρμογή (Kline, 2011; Wheaton et al., 1977) και οι τιμές των δεικτών CFI, TLI, NFI και RMSEA είναι επίσης αποδεκτές (Bentler & Bonett, 1980; Byrne, 1994; Fabrigar et al., 1999; Hu & Bentler, 1999; Schumacker & Lomax, 2004). Οι τιμές του μέτρου KMO (Hair et al., 2006; Kaiser, 1974) και του δείκτη σφαιρικότητας του Bartlett (Bartlett, 1951) δείχνουν επίσης ότι η δειγματοληψία είναι επαρκής και ότι η ανάλυση μπορεί να αποφέρει διακριτούς και αξιόπιστους παράγοντες. Το μοντέλο παρουσιάζει παρόμοια προσαρμογή και για τη Β΄ τάξη. Ωστόσο, το μοντέλο δεν παρουσίασε αντίστοιχα καλή προσαρμογή για τα δεδομένα που παρήχθησαν από τη Γ΄ τάξη, καθώς μόνο ο λόγος χ^2/df ήταν χαμηλότερος από την τιμή 5, που υποδεικνύει αποδεκτή προσαρμογή.

Πίνακας 2. Δείκτες προσαρμογής του μοντέλου ανά τάξη και συνολικά

	KMO	Δείκτης σφαιρικότητας του Bartlett	Δείκτες Προσαρμογής				
			χ^2/df	CFI	TLI	NFI	RMSEA
Α΄ Τάξη	0,84	χ^2 (300) = 1399, p <,001	1,870	0,954	0,949	0,907	0,074
Β΄ Τάξη	0,82	χ^2 (300) = 1408, p <,001	1,994	0,944	0,937	0,895	0,076
Γ΄ Τάξη	0,60	χ^2 (300) = 1580, p <,001	3,463	0,787	0,749	0,728	0,121
Σύνολο	0,82	χ^2 (300) = 4414, p <,001	4,914	0,934	0,920	0,918	0,087

Αποδεκτά όρια: KMO > 0,60 (Kaiser, 1974) ή 0,50 (Hair et al., 2006), Δείκτης σφαιρικότητας του Bartlett στατιστικά σημαντικός (Bartlett, 1951), χ2/df <3 (Kline, 2011) ή 5 (Wheaton et al., 1977), CFI < 0,95 (Hu & Bentler, 1999) ή 0,90 (Byrne, 1994), TLI > 0,90 (Bentler & Bonett, 1980) ή 0,95 (Hu & Bentler, 1999), NFI > 0,90 (Byrne, 1994) ή 0,95 (Schumacker & Lomax, 2004), 0,05> RMSEA > 0,1 (Fabrigar et al., 1999)

Τέλος, τα αποτελέσματα της επιβεβαιωτικής ανάλυσης παραγόντων με τα δεδομένα όλων των τάξεων του δείγματος έδειξαν αποδεκτές τιμές για όλους τους δείκτες προσαρμογής. Επίσης, φανερώνουν επαρκή δειγματοληψία με διακριτούς και αξιόπιστους παράγοντες.

Συμπερασματικά, το μοντέλο παρουσιάζει καλή προσαρμογή με τα δεδομένα όλων των τάξεων, αλλά καλύτερη για τις Α΄ και Β΄ τάξεις μεμονωμένα.

Χαρακτηριστικά ερωτημάτων

Δείκτης δυσκολίας

Η προσαρμοσμένη δοκιμασία BCTt περιλαμβάνει ερωτήσεις μικτής δυσκολίας για την Α΄ και Β΄ τάξη, καθώς η μέση τιμή της δυσκολίας της κλίμακας για καθεμία τάξη βρίσκεται πολύ κοντά στην τιμή 0,50, όπως προτείνεται (Azevedo et al., 2019). Οι ερωτήσεις παρουσιάζουν διαφοροποιούμενη δυσκολία και για τους/τις μαθητές/τριες της Γ΄ τάξης, αλλά η μέση τιμή του δείκτη δυσκολίας υπολογίστηκε σε 0,63, υποδηλώνοντας πιο εύκολη δοκιμασία σε σύγκριση με την Α΄ και Β΄ τάξη. Τέλος, η μέση τιμή του δείκτη δυσκολίας της δοκιμασίας για όλες τις τάξεις υπολογίστηκε σε 0,51, που φανερώνει ισορροπημένη κατανομή του επιπέδου δυσκολίας στο σύνολο της κλίμακας. Η κλίμακα, ακόμα, περιλαμβάνει στοιχεία με αυξανόμενη δυσκολία, ενώ η τιμή της αντίστοιχης γραμμής τάσης υποδηλώνει ότι η τελευταία μπορεί να ερμηνεύσει ένα σχετικά υψηλό ποσοστό (περίπου 73%) της διακύμανσης στις τιμές του δείκτη δυσκολίας και επομένως προσαρμόζεται καλά στα δεδομένα.

Δείκτης διακριτικής ικανότητας

Οι τιμές του δείκτη διακριτικής ικανότητας για τις ερωτήσεις στην υποομάδα της Α΄ τάξης έδειξαν ότι συνολικά η δοκιμασία είναι αποτελεσματική στη διάκριση μαθητών/τριών που έχουν υψηλή και χαμηλή επίδοση στη δοκιμασία, καθώς η μέση τιμή του δείκτη ήταν 0,53. Η μέση τιμή του δείκτη ήταν ικανοποιητική (0,51) και για την υποομάδα της Β΄ τάξης. Η μέση τιμή του δείκτη ήταν χαμηλότερη για την υποομάδα της Γ΄ τάξης (0,41), φανερώνοντας πως η δοκιμασία μπορεί να διακρίνει λιγότερο αποτελεσματικά μεταξύ μαθητών/τριών διαφορετικών επιδόσεων.

Συνολικά, η τιμή του δείκτη διακριτικής ικανότητας (0,53) έδειξε ότι οι ερωτήσεις της προσαρμοσμένης κλίμακας BCTt μπορούν επαρκώς να διακρίνουν τους/τις μαθητές/τριες διαφορετικών επιδόσεων όλων των τάξεων, αλλά έχουν μεγαλύτερη διακριτική ισχύ για την Α΄ και τη Β΄ τάξη. Δεδομένου ότι οι μεμονωμένες τιμές του δείκτη, δηλαδή οι τιμές για ξεχωριστές ερωτήσεις, πάντοτε ξεπέρασαν το όριο του 0,20 (El-Hamamsy et al., 2022), αποφασίσαμε να μην αφαιρέσουμε καμία ερώτηση από την κλίμακα.

Αξιοπιστία

Δείκτης εσωτερικής συνέπειας

Η αξιολόγηση της εσωτερικής συνέπειας της προσαρμοσμένης κλίμακας BCTt και των υποκλιμάκων της για καθεμία τάξη και για το σύνολο του δείγματος έδειξε ότι οι υποκλίμακες της *Ακολουθίας,* του *Απλού Βρόχου* και του *Φωλιασμένου Βρόχου* διαθέτουν επαρκή εσωτερική συνέπεια και ικανοποιητικού βαθμού συσχετίσεις μεταξύ των ερωτήσεων για όλες τις τάξεις του δείγματος. Στις υποκλίμακες *Αν-τότε, Αν-τότε-αλλιώς* και *Όσο* παρατηρήσαμε επαρκείς ή χαμηλές τιμές του δείκτη KR20 για όλες τις τάξεις. Ωστόσο, οι συγκεκριμένες υποκλίμακες περιείχαν λιγότερες ερωτήσεις και γι΄ αυτό εξετάστηκαν οι μέσες τιμές των ενδοστοιχειακών συσχετίσεων, οι οποίες βρέθηκαν εντός των προτεινόμενων τιμών.

Τέλος, η προσαρμοσμένη κλίμακα, αποτελούμενη από 25 ερωτήσεις, βρέθηκε πως διαθέτει στο σύνολο της υψηλές τιμές του δείκτη KR20 (0,80 ως 0,89) για τις τάξεις του δείγματος και στο σύνολό της (0,89) και επομένως αξιόλογη εσωτερική συνέπεια της κλίμακας (Cohen, 1977).

Μάλιστα, οι δείκτες δε βελτιώθηκαν κατά την αφαίρεση ερωτήσεων της κλίμακας και επομένως αποφασίσαμε να διατηρήσουμε όλα τα στοιχεία της πρωτότυπης κλίμακας.

Αξιοπιστία εξέτασης-επανεξέτασης

Η αξιοπιστία εξέτασης-επανεξέτασης ελέγχθηκε σε δείγμα 64 μαθητών/τριών από τις τρεις τάξεις του δείγματος σε δύο χρονικά σημεία που απείχαν μεταξύ τους 14 ως 18 ημέρες (Streiner et al., 2015). Για την Α' τάξη (N = 19), ο συντελεστής ICC υπολογίστηκε σε 0,804 (95% CI [0,559, 0,919], F (18, 18) = 9,183, p <,001), για τη Β' τάξη (N = 21) σε 0,815 (95% CI [0,598, 0,921], F (20, 20) = 9,809, p <,001) και για τη Γ' τάξη (N = 24) σε 0,728 (95% CI [0,466, 0,872], F (23, 23) = 6,353, p <,001). Ή τιμή του συντελεστή για τους/τις μαθητές/τριες όλων των τάξεων υπολογίστηκε σε 0,728 (95% CI [0,466, 0,872], F (23, 23) = 6,353, p <,001). Σε όλες τις περιπτώσεις υπάρχει υψηλός βαθμός αξιοπιστίας εξέτασης-επανεξέτασης.

Ανάπτυξη εννοιών ΥΣ μαθητών/τριών

Η μέση τιμή των βαθμολογιών των μαθητών/τριών όλων των τάξεων υπολογίστηκε σε 15,44 και η διάμεσός τους σε 16,00. Επιπλέον, η τιμή της τυπικής απόκλισης υπολογίστηκε σε 6,20 που υποδεικνύει μέτρια διασπορά των παρατηρήσεων γύρω από τη μέση τιμή τους. Υπολογίστηκε, ακόμα η βαθμολογία των μαθητών/τριών καθεμίας τάξης (βλ. Πίνακα 3).

Πίνακας 3. Περιγραφικά στατιστικά δεδομένα των βαθμολογιών ανά τάξη

Τάξη	*N*	Μ.Τ.	Διάμεσος	Τ.Α.	Ελάχιστο	Μέγιστο
Α'	160	12,44	13,00	6,22	0,00	24,00
Β'	172	13,49	14,00	5,96	0,00	24,00
Γ'	185	16,35	16,00	5,03	7,00	25,00
Δ'	156	19,58	21,00	5,07	6,00	25,00
Σύνολο	673	15,44	16,00	6,20	0,00	25,00

Συνολικά, τα περιγραφικά στατιστικά δεδομένα υποδεικνύουν αυξανόμενη επίδοση από τις μικρότερες προς τις μεγαλύτερες τάξεις. Οι μαθητές/τριες της Δ' τάξης σημείωσαν τη μεγαλύτερη μέση τιμή (Μ.Τ. = 19,58), ενώ 26 απ' αυτούς/ές (περίπου το 17% του συνόλου της τάξης) πέτυχαν τη μέγιστη βαθμολογία των 25 βαθμών, που μας οδήγησε στη διαπίστωση της ύπαρξης φαινομένου οροφής. Λαμβάνοντας υπόψη τον ευρέως υιοθετούμενο δείκτη του 15% του δείγματος που επιτυγχάνει τη μέγιστη βαθμολογία (Terwee et al., 2007), αποφασίσαμε τον αποκλεισμό των δεδομένων που αφορούν τη Δ' τάξη από περαιτέρω ανάλυση.

Διαφορές στον βαθμό ανάπτυξης της ΥΣ

Διαφορές ως προς την τάξη

Η μέση τιμή της βαθμολογίας ήταν διαφορετική σε στατιστικά σημαντικό βαθμό μεταξύ των μαθητών/τριών διαφορετικών τάξεων, Welch's F(3, 331,737) = 23,685, p <,001, αν και η επίδραση της τάξης φοίτησης στη συνολική βαθμολογία ήταν μέτρια, όπως υποδεικνύει η τιμή του συντελεστή η^2, που υπολογίστηκε σε 0,08. Η διενέργεια πολλαπλών συγκρίσεων των μέσων τιμών ανά ζεύγη με τη μέθοδο Games-Howell (Toothaker, 1993) έδειξε ότι στατιστικά σημαντική ήταν η αύξηση της μέσης τιμής της βαθμολογίας από την Α' ως τη Γ' τάξη (3,91, 95% CI [2,46, 5,36]), p <,001) αλλά και από τη Β' ως τη Γ' (2,86, 95% CI [1,48, 4,24], p <,001).

Διαφορές ως προς το φύλο

Η διενέργεια του στατιστικού ελέγχου *t*-test ανεξάρτητων δειγμάτων για τη διερεύνηση διαφορών στις βαθμολογίες της προσαρμοσμένης δοκιμασίας BCTt μεταξύ μαθητών (*M.T.* = 14,41, *T.A.* = 6,03) και μαθητριών (*M.T.* = 14,01, *T.A.* = 5,91) έδειξε ότι δεν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ τους, $t(515) = 0{,}755$, $p = {,}451$. Ακόμα, η τιμή του συντελεστή *d* του Cohen που υπολογίστηκε σε 0,17 έδειξε ότι το φύλο των μαθητών/τριών έχει μικρή επίδραση στον βαθμό ανάπτυξης των εννοιών της ΥΣ.

Αλληλεπίδραση τάξης-φύλου

Η διενέργεια της ανάλυσης διακύμανσης κατά δύο παράγοντες για τη διερεύνηση της αλληλεπίδρασης της τάξης και του φύλου των μαθητών/τριών στην ανάπτυξη εννοιών της ΥΣ έδειξε ότι η αλληλεπίδραση των ανεξάρτητων μεταβλητών στην εξαρτημένη δεν είναι στατιστικά σημαντική, $F(2, 511) = 0{,}008$, $p = {,}992$, $\eta_p^2 = 0{,}00$). Μάλιστα, η τιμή του συντελεστή του μεγέθους επίδρασης η_p^2 είναι ελάχιστη, φανερώνοντας ότι η αλληλεπίδραση των παραγόντων δεν έχει πρακτική σημασία στην ερμηνεία της διακύμανσης στις βαθμολογίες της δοκιμασίας μεταξύ των υποκειμένων της έρευνας διαφορετικών τάξεων και φύλου.

Συμπεράσματα

Τα αποτελέσματά μας παρείχαν ολοκληρωμένη κατανόηση της εγκυρότητας και αξιοπιστίας της προσαρμοσμένης στα ελληνικά κλίμακας BCTt για την αξιολόγηση της ανάπτυξης εννοιών ΥΣ από μαθητές δημοτικού σχολείου. Αρχικά, τα αποτελέσματα της ψυχομετρικής επικύρωσης της προσαρμοσμένης κλίμακας έδειξαν ότι διαθέτει ισχυρή εγκυρότητα περιεχομένου και εναρμονίζεται με το θεωρητικό της μοντέλο και κυρίως για την Α' και Β' τάξη. Η αξιοπιστία της είναι επίσης μεγάλη και κυρίως για τις δύο μικρότερες τάξεις. Τα αποτελέσματά μας είναι συνεπή με τα ευρήματα των Zapata-Cáceres et al. (2020, 2021), καθώς και οι δύο μελέτες υποδηλώνουν ότι η κλίμακα μπορεί να είναι λιγότερο απαιτητική για τους/τις μεγαλύτερους/ες μαθητές/τριες. Επιπλέον, αν και αμφότερες οι μελέτες εντόπισαν κατάλληλο βαθμό δυσκολίας της κλίμακας, τα ευρήματά μας έδειξαν υψηλότερο επίπεδο δυσκολίας και χαμηλότερες βαθμολογίες από εκείνες του δείγματος του πρωτότυπου εργαλείου. Οι παρατηρούμενες διαφορές υπογραμμίζουν την επίδραση πολιτιστικών, εκπαιδευτικών και περιβαλλοντικών παραγόντων στις βαθμολογίες των μαθητών/τριών, όπως και τη δυνατότητα περαιτέρω προσαρμογής του εργαλείου για τις μεγαλύτερες τάξεις.

Τα ευρήματά μας, ακόμα, περιλαμβάνουν διαφορές στη μέση τιμή των βαθμολογιών της κλίμακας αξιολόγησης των εννοιών της ΥΣ μεταξύ μαθητών/τριών διαφορετικών τάξεων. Ειδικότερα, φαίνεται πως καθώς οι μαθητές/τριες προχωρούν σε μεγαλύτερες τάξεις, αυξάνονται οι επιδόσεις τους στις έννοιες της ΥΣ, σύμφωνα με τα ευρήματα κι άλλων ερευνών (Jiang & Wong, 2021; Polat et al., 2021; Rijke et al., 2018), ενισχύοντας τη θέση ότι η γνωστική ωριμότητα μπορεί να συμβάλλει στη βελτίωση των δεξιοτήτων ΥΣ (Román-González et al., 2017). Δεδομένου ότι το Πρόγραμμα Σπουδών για τις ΤΠΕ στην ελληνική εκπαίδευση δίνει έμφαση σε πιο σύνθετες δεξιότητες στη Γ' τάξη (π.χ. δομή επιλογής, εκσφαλμάτωση) σε σύγκριση με τις βασικές δεξιότητες που διδάσκονται στην Α' και Β' (π.χ. βασικές αλγοριθμικές δεξιότητες, απλές αυτοματοποιημένες λύσεις) οι διαφορές μεταξύ των μαθητών/τριών των τάξεων μπορούν να αποδοθούν στην έκθεση σε διαφορετικής πολυπλοκότητας έννοιες, επιλεγμένες με βάση τα πορίσματα της αναπτυξιακής ψυχολογίας. Συνολικά, οι παρατηρούμενες διαφορές τονίζουν τη σημασία προσαρμογής των προσεγγίσεων για την καλλιέργεια και την αξιολόγηση της ΥΣ στο αναπτυξιακό στάδιο των μαθητών/τριών.

Στο πλαίσιο της διερεύνησης των διαφορών φύλου στον βαθμό ανάπτυξης των εννοιών της ΥΣ, δεν εντοπίσαμε στατιστικά σημαντικές διαφορές στις βαθμολογίες της κλίμακας

μεταξύ των δύο φύλων, όπως και άλλοι ερευνητές με δείγμα μικρότερων μαθητών/τριών (Angeli & Georgiou, 2023; Angeli & Valanides, 2020). Ωστόσο, τα ευρήματά μας αντιτίθενται στα ευρήματα ερευνών που εντόπισαν διαφορές φύλου, αλλά είτε το μέγεθος της επίδρασης ήταν μικρό (Polat et al., 2021; Román-González et al., 2017) είτε οι διαφορές αφορούσαν συγκεκριμένες δεξιότητες ΥΣ (Sullivan & Bers, 2013, 2016). Συνολικά, τα ευρήματά μας συμβάλλουν στη συνεχιζόμενη συζήτηση για τις διαφορές μεταξύ μαθητών/τριών δημοτικού σχολείου στις δεξιότητες ΥΣ. Τέλος, η έρευνά μας συμβάλλει στην περιορισμένη έρευνα για την αλληλεπίδραση τάξης-φύλου στην ανάπτυξη δεξιοτήτων ΥΣ. Τα αποτελέσματά μας έδειξαν πως δεν υπάρχει στατιστικά σημαντική αλληλεπίδραση μεταξύ τάξης και φύλου στην ανάπτυξη δεξιοτήτων ΥΣ, εν μέρει αντιτιθέμενα με εκείνα των Rijke et al. (2018), οι οποίοι εντόπισαν διαφορές φύλου που σχετίζονται με την ηλικία σε συγκεκριμένες δεξιότητες ΥΣ μετά την ηλικία των 9,5 ετών.

Τέλος, το γεγονός ότι η κλίμακα αποτελείται από δυαδικά στοιχεία ενδέχεται να την καθιστά λιγότερο αποτελεσματική στην ανίχνευση λεπτών διαφορών στην ανάπτυξη των εννοιών της ΥΣ. Σε μελλοντικές μελέτες θα μπορούσαν να προστεθούν ερωτήσεις πολλαπλής επιλογής ή δραστηριότητες ανοικτού τύπου. Επιπρόσθετα, το επίπεδο ανάπτυξης των εννοιών και δεξιοτήτων ΥΣ μπορεί να μην αποτυπώνεται επαρκώς όταν αξιολογείται μόνο με χρήση γραπτής δοκιμασίας. Περαιτέρω βελτίωση των στοιχείων και υποκλιμάκων βάσει των ευρημάτων της ψυχομετρικής επικύρωσης και της διερεύνησης των διαφορών τάξης και φύλου μπορεί να βελτιώσει τη συνολική ψυχομετρική ποιότητα και χρησιμότητα της κλίμακας σε εκπαιδευτικά και ερευνητικά πλαίσια.

Διαθεσιμότητα ερευνητικού εργαλείου

Η μετάφραση, πολιτισμική προσαρμογή και ψυχομετρική επικύρωση του BCTt πραγματοποιήθηκε στο πλαίσιο μεταδιδακτορικής έρευνας που εκπονήθηκε στο Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης της Σχολής Ανθρωπιστικών και Κοινωνικών Επιστημών του Πανεπιστημίου Θεσσαλίας. Το εργαλείο είναι ελεύθερα διαθέσιμο για ερευνητικούς σκοπούς στην ιστοσελίδα https://vourletsis.users.uth.gr/ .

Βιβλιογραφικές Αναφορές

Adams, C., Cutumisu, M., & Lu, C. (2019). Measuring K-12 Computational Thinking Concepts, Practices and Perspectives: An Examination of Current CT Assessments. In K. Graziano (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 275–285). Las Vegas: Association for the Advancement of Computing in Education (AACE). Retrieved May 26, 2024, from https://www.learntechlib.org/primary/p/207654/

Angeli, C., & Georgiou, K. (2023). Investigating the effects of gender and scaffolding in developing preschool children's computational thinking during problem-solving with Bee-Bots. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.757627

Angeli, C., & Giannakos, M. (2020). Computational thinking education: Issues and challenges. *Computers in Human Behavior*, *105*. https://doi.org/10.1016/j.chb.2019.106185

Angeli, C., & Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. *Computers in human behavior*, *105*, 105954. https://doi.org/10.1016/j.chb.2019.03.018

Annamalai, S., Che Omar, A., & Abdul Salam, S. N. (2022). Review of Computational Thinking Models in Various Learning Fields. *International Journal of Education, Psychology and Counseling*, *7*(48), 562–574. https://doi.org/10.35631/ijepc.748042

Arafat, S., Chowdhury, H., Qusar, M., & Hafez, M. (2016). Cross Cultural Adaptation and Psychometric Validation of Research Instruments: a Methodological Review. *Journal of Behavioral Health*, *5*(3), 129. https://doi.org/10.5455/jbh.20160615121755

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems*, *75*, 661–670. https://doi.org/10.1016/j.robot.2015.10.008

Azevedo, J. M., Oliveira, E. P., & Beites, P. D. (2019). Using Learning Analytics to evaluate the quality of multiple-choice questions. *The International Journal of Information and Learning Technology*, *36*(4), 322–341. https://doi.org/10.1108/ijilt-02-2019-0023

Babazadeh, M., & Negrini, L. (2022). How is computational thinking assessed in European K-12 education? A systematic review. *International Journal of Computer Science Education in Schools*, *5*(4), 3–19. https://doi.org/10.21585/ijcses.v5i4.138

Barnard, J. J. (1999). Item Analysis in Test Construction. *Advances in Measurement in Educational Research and Assessment*, 195–206. https://doi.org/10.1016/b978-008043348-6/50016-4

Bartlett, M. S. (1951). The Effect of Standardization on a χ^2 Approximation in Factor Analysis. *Biometrika*, *38*(3/4), 337–344. https://doi.org/10.2307/2332580

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, *6*. https://doi.org/10.3389/fpubh.2018.00149

Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia (Ribeirão Preto)*, *22*, 423–432. https://doi.org/10.1590/1982-43272253201314

Brennan K, & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the annual American educational research association meeting* (pp. 1–25). Retrieved May 26, 2024 from https://web.media.mit.edu/~kbrennan/files/Brennan_Resnick_AERA2012_CT.pdf

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*(1), 106–148. https://doi.org/10.1111/j.1467-6494.1986.tb00391.x

Byrne, B.M. (1994). *Structural Equation Modelling with EQS and EQS/Windows: Basic Concepts, Applications, and Programming*. Sage.

Cheung, A.K.L. (2014). Probability Proportional Sampling. In A.C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 5069–5071). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_2269

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Lawrence Erlbaum Associates, Inc.

Cutumisu, M., Adams, C., & Lu, C. (2019). A Scoping Review of Empirical Research on Recent Computational Thinking Assessments. *Journal of Science Education and Technology*, *28*(6), 651–676. https://doi.org/10.1007/s10956-019-09799-3

El-Hamamsy, L., Zapata-Cáceres, M., Marcelino, P., Bruno, B., Dehler Zufferey, J., Martín-Barroso, E., & Román-González, M. (2022). Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt). *Frontiers in Psychology*, *13*, 1082659. https://doi.org/10.3389/fpsyg.2022.1082659

Fabrigar, L. R., MacCallum, R. C., Wegener, D. T., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Flórez, F. B., Casallas, R., Hernández, M., Reyes, A., Restrepo, S., & Danies, G. (2017). Changing a Generation's Way of Thinking: Teaching Computational Thinking Through Programming. *Review of Educational Research*, *87*(4), 834–860. https://doi.org/10.3102/0034654317710096

Grover, S., & Pea, R. (2018). Computational Thinking: A competency whose time has come. In S. Sentence, E. Barendsen, & C. Schulte (Eds.), *Computer Science Education: Perspectives on teaching and learning in School (*pp. 19-38*)*. https://doi.org/10.5040/9781350057142.ch-003

Hazzan, O., Ragonis, N., & Lapidot, T. (2020). Computational thinking. In O. Hazzan, N. Ragonis, & T. Lapidot (Eds.), *Guide to Teaching Computer Science* (pp. 57–74). Springer. https://doi.org/10.1007/978-3-030-39360-1_4

Hingorjo, MR., Jaleel, F. (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *Journal of the Pakistan Medical Association, 62*(2), 142–7.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. *Journal of Organizational Behavior*, *18*(6), 667–683. https://doi.org/10.1155/2016/2696019

Jiang, S., & Wong, G. K. W. (2021). Exploring age and gender differences of computational thinkers in primary school: A developmental perspective. *Journal of Computer Assisted Learning*, *38*(1), 60–75. Portico. https://doi.org/10.1111/jcal.12591

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kim, H. S., Kim, S., Na, W., & Lee, W. J. (2021). Extending Computational Thinking into Information and Communication Technology Literacy Measurement. *ACM Transactions on Computing Education*, *21*(1), 1–25. https://doi.org/10.1145/3427596

Kim, H. Y. (2014). Statistical notes for clinical researchers: Two-way analysis of variance (ANOVA)-exploring possible interaction between factors. *Restorative Dentistry & Endodontics*, *39*(2), 143. https://doi.org/10.5395/rde.2014.39.2.143

Kline, R. B. (2011). *Principles and practice of structural equation modeling, Structural Equation Modeling*. Guilford Press.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160. https://doi.org/10.1007/BF02288391

Lu, C., Macdonald, R., Odell, B., Kokhan, V., Demmans Epp, C., & Cutumisu, M. (2022). A scoping review of computational thinking assessments in higher education. *Journal of Computing in Higher Education*, *34*(2), 416-461. https://doi.org/10.1007/s12528-021-09305-y

Lynn, M. R. (1986). Determination and Quantification Of Content Validity. *Nursing Research*, *35*(6), 382-386. https://doi.org/10.1097/00006199-198611000-00017

Mouza, C., Pan, Y.-C., Yang, H., & Pollock, L. (2020). A Multiyear Investigation of Student Computational Thinking Concepts, Practices, and Perspectives in an After-School Computing Program. *Journal of Educational Computing Research*, *58*(5), 1029–1056. https://doi.org/10.1177/0735633120905605

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books: New York.

Passos, M. P. V. D., Almeida, J. R., Santos, Y. H. S., Junior, E. P. P., Flores-Quispe, M. del P., Aquino, R.

Martufi, V., Barreto, M., & Amorim, L. D. A. F. (2023). Measurement Models with Binary Indicators: A Tutorial for the Assessment of Antenatal Care Quality. https://doi.org/10.21203/rs.3.rs-2860527/v1

Polat, E., Hopcan, S., Kucuk, S., & Sisman, B. (2021). A comprehensive assessment of secondary school students' computational thinking skills. *British Journal of Educational Technology*, *52*(5), 1965–1980. https://doi.org/10.1111/bjet.13092

Polya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400828678

Poulakis, E., & Politis, P. (2021). Computational Thinking Assessment: Literature Review. In T. Tsiatsos, S. Demetriadis, A. Mikropoulos, & V. Dagdilelis (Eds.), *Research on E-Learning and ICT in Education (*pp. 111–128*)*. https://doi.org/10.1007/978-3-030-64363-8_7

Rasch, D., Kubinger, K. D., & Moder, K. (2011). The two-sample t test: Pre-testing its assumptions does not pay off. *Statistical Papers, 52*, 219–231. https://doi.org/10.1007/s00362-009-0224-x

Rijke, W. J., Bollen, L., Eysink, T. H. S., & Tolboom, J. L. J. (2018). Computational Thinking in Primary School: An Examination of Abstraction and Decomposition in Different Age Groups. *Informatics in Education*, *17*(1), 77–92. https://doi.org/10.15388/infedu.2018.05

Román-González, M. (2015). Computational thinking test: Design guidelines and content validation. In *EDULEARN15 Proceedings* (pp. 2436–2444). IATED. Retrieved May 26, 2024, from https://library.iated.org/view/ROMANGONZALEZ2015COM

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining Assessment Tools for a Comprehensive Evaluation of Computational Thinking Interventions. *Computational Thinking Education*, 79–98. https://doi.org/10.1007/978-981-13-6528-7_6

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior*, *72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Journal of Central South University. Medical sciences*, *37*(2), 152–155. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

Singh, S. (2003). Simple Random Sampling. In *Advanced Sampling Theory with Applications* (pp. 71–136). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0789-4_2

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA.

Sullivan, A., & Bers, M.U. (2013). Gender differences in kindergarteners' robotics and programming achievement. *International Journal of Technology and Design Education, 23*, 691–702. https://doi.org/10.1007/s10798-012-9210-z

Sullivan, A., & Bers, M.U. (2016). Girls, Boys, and Bots: Gender Differences in Young Children's Performance on Robotics and Programming Tasks. *Journal of Information Technology Education: Innovations in Practice*, *15*, 145–165. https://doi.org/10.28945/3547

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, 103798. https://doi.org/10.1016/j.compedu.2019.103798

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Toothaker, L. (1993). *Multiple Comparison Procedures*. Sage. https://doi.org/10.4135/9781412985178

Tsang, S., Royse, C., & Terkawi, A. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia*, *11*(5), 80. https://doi.org/10.4103/sja.sja_203_17

Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology*, *8*, 84–136. https://doi.org/10.2307/270754

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2011). Research notebook: Computational thinking—What and why? *The link magazine.* Retrieved May 26, 2024, from https://www.cs.cmu.edu/link/research-notebook-computational-thinking-what-and-why.

Zapata-Cáceres, M. Martín-Barroso, E., & Román-González, M. (2021). BCTt: Beginners Computational Thinking Test. In *Understanding computing education (Vol 1). Proceedings of the Raspberry Pi Foundation Research Seminar series*. Retrieved May, 26, 2024, from www.rpf.io/seminar-proceedings-2020

Zapata-Cáceres, M., Marcelino, P., El-Hamamsy, L., & Martín-Barroso, E. (2024). A Bebras Computational Thinking (ABC-Thinking) program for primary school: Evaluation using the competent computational thinking test. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12441-w

Zapata-Cáceres, M., Martín-Barroso, E., & Román-González, M. (2020). Computational Thinking Test for Beginners: Design and Content Validation. In *2020 IEEE Global Engineering Education Conference (EDUCON)*. https://doi.org/10.1109/educon45650.2020.9125368