

THÍCH NGHI VÀ HIỆU CHUẨN TRẮC NGHIỆM SÀNG LỌC RỐI LOẠN TÍNH TOÁN TTA TRÊN HỌC SINH TỪ LỚP 1 ĐẾN LỚP 6 TẠI THÀNH PHỐ HỒ CHÍ MINH

ADAPTATION AND VALIDATION OF THE TEMPOTEST AUTOMATISEREN (TTA) SCREENING TEST
FOR DYSCALCULIA IN VIETNAMESE GRADE 1-6 STUDENTS IN HO CHI MINH CITY

Lê Thị Mai Liên

Trường Đại học Khoa học Xã hội và Nhân văn - Đại học Quốc gia Thành phố
Hồ Chí Minh
Email: mailien.lethi@hcmussh.edu.vn

Article history

Received: 26/6/2026

Accepted: 27/7/2026

Published: 20/9/2026

Keywords

Screening test, percentile norms, dyscalculia, primary school students, TempoTest Automatiseren (TTA)

ABSTRACT

Dyscalculia is a specific learning disorder of numerical and arithmetic skills, affecting approximately 3-6% of the general population (DSM-5; American Psychiatric Association, 2013). In the screening and diagnosis of dyscalculia in Vietnam, a notable gap is the absence of quantitative reference norms standardized on Vietnamese children, even for screening tools. This study presents the first results of adapting and calibrating the TempoTest Automatiseren (TTA; de Vos, 2010) screening test on a sample of 1,342 Grade 1-6 students in Ho Chi Minh City. Using descriptive statistics and percentile tables for each arithmetic operation and grade level, the study establishes the first percentile reference norms for the Vietnamese version of the TTA, and discusses how clinical percentile thresholds (Pc16, corresponding to -1 standard deviation per the INAMI regulation, with Pc10 and Pc25 as reference markers) can be used to identify students at risk of arithmetic difficulties. The results also highlight key cautions for interpretation, including ceiling effects on easier operations and some between-grade anomalies in performance. These norms are an initial step and require expansion and full psychometric validation before broad application.

1. Mở đầu

Rối loạn tính toán (dyscalculia) là một rối loạn chuyên biệt về học tập liên quan đến các kỹ năng nền tảng về số học, ảnh hưởng đến khoảng 3-6% dân số nói chung (American Psychiatric Association, 2013). Theo DSM-5, rối loạn tính toán được chẩn đoán khi thành tích số học (đo bằng trắc nghiệm chuẩn hóa) thấp hơn rõ rệt so với mức kỳ vọng theo độ tuổi, tồn tại dai dẳng và không được giải thích bởi khiếm khuyết trí tuệ, giác quan hay điều kiện học tập. Việc nhận diện trẻ có rối loạn tính toán là một tiến trình nhiều bước, đi từ sàng lọc đến chẩn đoán chuyên sâu, nhằm phân biệt trẻ chậm phát triển tính toán đơn thuần với trẻ thực sự có rối loạn, từ đó đề xuất can thiệp có chứng cứ phù hợp (Cohen Kadosh và cộng sự, 2013; Cook và cộng sự, 2013; Lafay và cộng sự, 2014; Noël, 2005).

Tại Việt Nam, lĩnh vực rối loạn chuyên biệt học tập mới được quan tâm trong khoảng một thập niên gần đây, với một khoảng trống lớn là sự thiếu hụt các công cụ sàng lọc, chẩn đoán được chuẩn hóa cho lượng giá xử lý số học. Tổng quan trước đó của tác giả về các công cụ lâm sàng phổ biến gồm công cụ sàng lọc (TTA, KRT-R) và các trắc nghiệm chẩn đoán chuẩn hóa (Tedi-Math, UDN-II, ZAREKI-R), điều này cho thấy ngay cả với một công cụ sàng lọc đơn giản, việc diễn giải điểm số phụ thuộc hoàn toàn vào một cơ sở tham chiếu định lượng phù hợp với dân số trẻ em Việt Nam, cơ sở mà hiện vẫn còn thiếu (Lê Thị Mai Liên, 2018).

Tính lưu loát tính toán (arithmetic fluency) là khả năng thực hiện các phép tính cơ bản một cách nhanh, chính xác và ít tốn công sức nhận thức. Trong phát triển bình thường, trẻ chuyển dần từ các chiến lược đếm ngón tay (finger counting) sang truy xuất trực tiếp các “sự kiện số học” (arithmetic facts) từ trí nhớ dài hạn (Geary, 2004). Trẻ có rối loạn tính toán đặc trưng bởi khiếm khuyết dai dẳng trong truy xuất sự kiện số học và sự chậm trễ trong tự động hóa (Geary và cộng sự, 2012). Vì thiếu hụt về lưu loát là một trong những biểu hiện sớm và bền vững nhất của khó khăn tính toán, việc đo lường tính lưu loát trong một khoảng thời gian giới hạn trở thành một chỉ báo nhạy và ít tốn chi phí để sàng lọc nguy cơ.

TTA (TempoTest Automatiseren) là một công cụ sàng lọc điển hình đo tính lưu loát tính toán: HS thực hiện càng nhiều phép tính đúng càng tốt trong thời gian quy định. Công cụ gồm bốn tiểu trắc nghiệm (cộng, trừ, nhân, chia),

mỗi tiểu trắc nghiệm 50 mục thực hiện trong 2 phút, được chuẩn hóa và sử dụng rộng rãi ở Hà Lan và vùng Flanders (Bi) (de Vos, 2010). So sánh với các công cụ cùng lĩnh vực cho thấy rõ lý do TTA được chọn cho việc nghiên cứu thích nghi tại Việt Nam. Các trắc nghiệm chẩn đoán chuyên sâu như Tedi-Math hay ZAREKI-R bao phủ nhiều thành phần nhận thức số nhưng phải thực hiện cá nhân, kéo dài 45-60 phút và đòi hỏi người đánh giá được đào tạo chuyên biệt - điều kiện chưa sẵn có ở phần lớn trường học Việt Nam. Ngược lại, các trắc nghiệm đa thành phần thể hệ mới như MathPro Test (Karagiannakis và Noël, 2020) tuy thực hiện được theo nhóm nhưng phụ thuộc hạ tầng máy tính và chứa nhiều nhiệm vụ có yếu tố ngôn ngữ, khiến quy trình thích nghi xuyên ngôn ngữ phức tạp hơn. TTA nằm ở vị trí trung gian có lợi: nội dung mục thuần túy là kí hiệu số nên gần như trung tính về ngôn ngữ và văn hóa, chỉ 8 phút cho cả bốn phép tính, thực hiện theo lớp bằng giấy - bút, chi phí thấp, trong khi vẫn đo đúng chỉ báo sớm và bền vững nhất của khó khăn tính toán. TTA do đó không thay thế chẩn đoán chuyên sâu mà là bước sàng lọc đầu tiên khả thi nhất về nguồn lực. Do điểm thô của một công cụ sàng lọc tự nó không mang ý nghĩa lâm sàng, TTA đã được thích nghi ở nhiều bối cảnh ngôn ngữ: Lafay và cộng sự (2015) chuẩn hóa cho trẻ em nói tiếng Pháp ở vùng Quebec và nhận thấy trẻ Québec đạt điểm thấp hơn trẻ Hà Lan, minh chứng cho thấy chuẩn tham chiếu cần được xây dựng riêng cho từng bối cảnh. Ở châu Á, việc sàng lọc phần lớn vẫn dựa trên các bài kiểm tra thành tích học đường; tại Trung Quốc, Wang và cộng sự (2024) tổng hợp 54 nghiên cứu (34.815 trẻ) cho thấy tỉ lệ khó khăn toán dao động theo điểm cắt (khoảng 10,8% ở ngưỡng -1 độ lệch chuẩn và 5,8% ở -1,5 độ lệch chuẩn). Các công cụ đo lưu loạt tính toán đơn giản, được chuẩn hóa như TTA còn ít hiện diện trong khu vực.

Về ngưỡng sàng lọc, trong lĩnh vực nhận thức số, một trẻ thường được xem là có khó khăn khi điểm số thấp hơn 1 độ lệch chuẩn so với thành tích mong đợi theo độ tuổi học đường - ngưỡng nghiêm ngặt hơn mức 2 độ lệch chuẩn thường dùng trong các lĩnh vực nhận thức nói chung (Institut National d'assurance maladie-invalidité, 2024). Trên thang bách phân vị, mức -1 độ lệch chuẩn tương ứng xấp xỉ bách phân vị thứ 16 (Pc16). Bên cạnh Pc16, thực hành lâm sàng còn sử dụng Pc10 (chặt hơn, độ đặc hiệu cao, ít dương tính giả) và Pc25 (rộng hơn, độ nhạy cao, ít bỏ sót) để phân tầng nguy cơ, cũng như các ngưỡng nghiêm ngặt hơn (-1,5 độ lệch chuẩn $\approx$ Pc7, hoặc $Pc \leq 5$) cho các trường hợp điển hình; DSM-5 không quy định một ngưỡng cứng mà yêu cầu thành tích "thấp hơn đáng kể" kèm tính dai dẳng (Murphy và cộng sự, 2007). Việc lựa chọn điểm cắt tùy thuộc mục đích sàng lọc và nguồn lực can thiệp sẵn có.

Trên cơ sở đó, bài báo trình bày kết quả bước đầu thích nghi và hiệu chuẩn trắc nghiệm sàng lọc TTA trên HS Việt Nam, nhằm khảo sát một số đặc tính tâm trắc cơ bản và thiết lập bộ chuẩn tham chiếu bách phân vị đầu tiên cho công cụ này. Bài báo tập trung trả lời 02 câu hỏi tương ứng với 02 mục tiêu: *Thứ nhất*, đánh giá chất lượng tâm trắc của phiên bản TTA tiếng Việt thông qua 03 tiêu chí: (1) độ tin cậy nhất quán nội tại; (2) chất lượng các mục trắc nghiệm về độ khó và độ phân biệt; (3) hiệu lực phát triển, thể hiện ở khả năng phản ánh sự gia tăng tính lưu loạt tính toán theo cấp lớp; *Thứ hai*, xây dựng bộ chuẩn tham chiếu bằng cách xác lập phân bố điểm bách phân vị của HS lớp 1 đến lớp 6 theo từng phép tính, đồng thời làm rõ ý nghĩa của việc đối chiếu kết quả của một HS với các ngưỡng tham chiếu lâm sàng (Pc16 theo INAMI, cùng Pc10, Pc25 và Pc2) cũng như những giới hạn cần lưu ý khi diễn giải kết quả. Hai câu hỏi nghiên cứu có mối quan hệ kế tiếp: chỉ khi phiên bản TTA tiếng Việt đáp ứng các yêu cầu tâm trắc tối thiểu của một công cụ sàng lọc thì bộ chuẩn tham chiếu được xây dựng mới có cơ sở để sử dụng trong thực tiễn.

2. Phương pháp nghiên cứu

Trắc nghiệm TTA (de Vos, 2010) yêu cầu HS thực hiện nhanh và chính xác nhất có thể các phép tính cộng, trừ, nhân, chia trên 50 mục cho mỗi phép trong thời gian 2 phút; điểm số là số câu trả lời đúng, phản ánh mức độ tự động hóa tính toán. Kết quả TTA chỉ có ý nghĩa phát hiện nguy cơ, chưa đủ để kết luận trẻ có rối loạn tính toán. Quá trình thích nghi TTA tại Việt Nam được thực hiện trong khuôn khổ dự án đào tạo "Nhận diện, lượng giá và can thiệp rối loạn chuyên biệt học tập trong bối cảnh học đường" (Trường Đại học Khoa học Xã hội và Nhân văn - UCLouvain - ULiège), từ tháng 5/2018 đến tháng 11/2021. Về quy trình thích nghi, cần lưu ý trước một đặc điểm của TTA: toàn bộ 200 mục chỉ gồm biểu thức số học viết bằng kí hiệu số Ả Rập (ví dụ $7 + 5 =$, $42 : 6 =$), không chứa từ ngữ và không có bài toán có lời văn. Khối lượng nội dung cần chuyển ngữ vì thế là tối thiểu và nguy cơ lệch nghĩa xuyên ngôn ngữ ở cấp độ mục không đáng kể; công việc thích nghi tập trung vào phần hướng dẫn và điều kiện thực hiện, gồm 04 bước: (1) Phân chỉ dẫn cho nghiệm viên, cho HS và các mục ví dụ được dịch xuôi sang tiếng Việt bởi thành viên nhóm nghiên cứu thông thạo cả hai ngôn ngữ, rồi dịch ngược bởi một người dịch độc lập không tiếp cận bản gốc; hai bản được đối chiếu và thống nhất; (2) Bản tiếng Việt được nhóm chuyên gia gồm giảng viên tâm lý học phát triển, chuyên viên tâm lý học đường và GV tiểu học rà soát về tính rõ ràng và sự tương thích với cách trình bày phép tính trong chương trình môn Toán hiện hành; (3) Cấu trúc bài trắc nghiệm được giữ nguyên hoàn toàn so với bản gốc - số mục, trật tự mục theo độ khó tăng dần, giới hạn 2 phút cho mỗi tiểu trắc nghiệm - nhằm bảo toàn khả năng

so sánh xuyên bối cảnh; (4) Bản thích nghi được thử nghiệm trên nhóm nhỏ HS thuộc nhiều khối lớp để kiểm tra tính khả thi của quy trình thực hiện tập thể, thời lượng và mức độ hiệu chỉ dẫn; phản hồi được dùng để hoàn thiện phân hướng dẫn trước khi thu thập dữ liệu chính thức.

Mẫu hiệu chuẩn gồm 1342 HS từ lớp 1 đến lớp 6 tại TP. Hồ Chí Minh (tuổi trung bình toàn mẫu 116 tháng), phân bố theo lớp: Lớp 1 (n = 190), Lớp 2 (n = 110), Lớp 3 (n = 214), Lớp 4 (n = 330), Lớp 5 (n = 233) và Lớp 6 (n = 265). Về giới tính: 599 nam (44,6%), 555 nữ (41,4%) và 188 trường hợp thiếu thông tin về giới tính (14,0%). Theo thiết kế gốc của TTA, HS lớp 1 và lớp 2 chỉ thực hiện phép cộng và phép trừ; từ lớp 3 trở lên thực hiện đủ cả bốn phép tính. Bài báo sử dụng phương pháp chọn mẫu thuận tiện theo cụm: đơn vị chọn mẫu là lớp học nguyên vẹn tại các trường công lập đồng ý tham gia dự án trên địa bàn TP. Hồ Chí Minh; trong mỗi trường, các lớp được chọn để bao phủ đủ các khối lớp mục tiêu và toàn bộ HS có mặt đều được mời tham gia. Cách chọn mẫu này phù hợp với mục tiêu thu thập cỡ mẫu lớn cho bước hiệu chuẩn ban đầu và với hình thức thực hiện tập thể của TTA, song đặt ra giới hạn về tính đại diện: mẫu không được phân tầng theo quận/huyện, loại hình trường hay điều kiện KT-XH của gia đình và số lượng HS không đồng đều giữa các khối (từ n = 110 ở lớp 2 đến n = 330 ở lớp 4). Mẫu vì vậy nên được hiểu là đại diện cho HS tại các trường tham gia dự án ở TP. Hồ Chí Minh, chưa phải mẫu đại diện xác suất cho toàn thành phố hay cho cả nước. Về 188 trường hợp thiếu thông tin giới tính (14,0%), đây là dữ liệu khuyết ở biến nhân khẩu học chứ không phải ở biến kết quả: ở một số lớp, phiếu trả lời được thu ẩn danh hoàn toàn và ô giới tính do HS tự điền đã bị bỏ trống hoặc điền không hợp lệ. Toàn bộ 1342 HS đều có đầy đủ dữ liệu điểm TTA và cấp lớp - hai biến duy nhất được dùng để lập bảng chuẩn bách phân vị. Vì bộ chuẩn được xây dựng theo cấp lớp chứ không theo giới tính, tình trạng khuyết dữ liệu này không làm thay đổi cỡ mẫu hay kết quả của bất kỳ phân tích nào được báo cáo; ảnh hưởng thực sự là mất khả năng kiểm định khác biệt theo giới, đây là lý do tác giả không đề xuất chuẩn riêng theo giới. Bài báo được thực hiện tuân thủ các nguyên tắc đạo đức trong nghiên cứu giáo dục. Do dữ liệu được thu thập theo hình thức tập thể trong giờ học chính khóa và chỉ bao gồm bài đánh giá năng lực tính toán (4 bài test, mỗi bài 2 phút, tổng thời lượng 8 phút) với mức rủi ro tối thiểu, tác giả đã nhận được sự đồng thuận của Ban Giám hiệu nhà trường cùng GV phụ trách lớp. Nghiên cứu không thu thập bất kỳ thông tin định danh cá nhân nào của HS; toàn bộ dữ liệu được ẩn danh và chỉ được phân tích, báo cáo ở cấp độ nhóm. Việc tham gia là tự nguyện và không ảnh hưởng đến kết quả học tập của HS.

Dữ liệu được xử lý bằng phần mềm SPSS. Chiến lược phân tích bám sát 02 câu hỏi nghiên cứu. Để trả lời câu hỏi thứ nhất, 03 nhóm phân tích được thực hiện tương ứng với 03 tiêu chí đã nêu. Độ tin cậy được đánh giá qua hệ số nhất quán nội tại Cronbach's Alpha trên dữ liệu nhị phân ở cấp độ từng mục (1 = đúng, 0 = sai; mục chưa kịp thực hiện trong thời gian quy định được tính là chưa đạt). Đặc trưng của các mục được khảo sát qua độ khó (p) và độ phân biệt (hệ số tương quan mục-tổng hiệu chỉnh, rit). Hiệu lực phát triển được kiểm định bằng phân tích phương sai một yếu tố (ANOVA) so sánh điểm theo cấp lớp, kèm hệ số eta bình phương (η^2) để ước lượng mức tác động và hệ số tương quan Pearson giữa điểm số với cấp lớp và với tuổi tính theo tháng. Để trả lời câu hỏi thứ hai về chuẩn tham chiếu, với mỗi phép tính và cấp lớp, nghiên cứu lập các chỉ số thống kê mô tả (trung bình, trung vị, độ lệch chuẩn) và bảng bách phân vị (Pc5, Pc10, Pc16, Pc25, Pc50, Pc75). Mốc Pc16 (≈ -1 độ lệch chuẩn) được dùng làm ngưỡng lâm sàng chính theo INAMI; Pc10 và Pc25 là các mốc tham chiếu hỗ trợ. Kết quả của nhóm phân tích thứ nhất được trình bày ở mục 3.1-3.2 (bảng 1-3), của nhóm thứ hai ở mục 3.3 (bảng 4). Nghiên cứu dừng ở bước thiết lập chuẩn tham chiếu định lượng ban đầu; việc xác định điểm cắt tối ưu dựa trên một nhóm lâm sàng đã được chẩn đoán nằm ngoài phạm vi dữ liệu hiện có.

3. Kết quả nghiên cứu

3.1. Độ tin cậy và đặc trưng của các mục

Kết quả được trình bày theo đúng trật tự 02 câu hỏi: mục 3.1 và 3.2 kiểm định ba tiêu chí tâm trắc của câu hỏi thứ nhất, mục 3.3 xây dựng bộ chuẩn của câu hỏi thứ hai. Xét tiêu chí thứ nhất là độ tin cậy, trên toàn mẫu, cả bốn tiểu trắc nghiệm đều có độ tin cậy rất cao, với hệ số Cronbach's Alpha dao động từ 0,959 (phép nhân) đến 0,967 (phép cộng); khi gộp toàn bộ 200 mục, hệ số alpha của toàn thang đạt 0,983, vượt xa ngưỡng 0,70 thường được khuyến nghị (bảng 1).

Bảng 1. Hệ số Cronbach's Alpha và thống kê mô tả của các tiểu trắc nghiệm TTA trên toàn mẫu

Tiểu trắc nghiệm	Số mục	N	Cronbach's α	ĐTB $\pm$ ĐLC
Phép cộng	50	1.342	0,967	29,61 $\pm$ 11,68
Phép trừ	50	1.342	0,963	26,49 $\pm$ 11,25
Phép nhân	50	1.040	0,959	37,19 $\pm$ 10,43
Phép chia	50	1.042	0,966	36,02 $\pm$ 11,78
Toàn thang (4 tiểu trắc nghiệm)	200	1.040	0,983	—

Ghi chú: ĐTB = điểm trung bình (số câu đúng); ĐLC = độ lệch chuẩn. Phép nhân và phép chia chỉ được thực hiện từ lớp 3 trở lên.

Xét tiêu chí thứ hai là chất lượng của các mục (bảng 2), độ khó trải rộng từ rất dễ đến khó, cho phép phân hóa tốt năng lực HS ở nhiều mức độ. Độ phân biệt trung bình của cả bốn tiểu trắc nghiệm đều ở mức cao (rit trung bình từ 0,50 đến 0,55). Một số ít mục có độ phân biệt thấp (rit < 0,20) đều là các mục “khởi động” đầu mỗi tiểu trắc nghiệm có độ khó gần như bằng trần ($p > 0,95$) phù hợp với thiết kế của một trắc nghiệm tính nhanh và không làm giảm chất lượng đo lường tổng thể. Độ khó của mục tương quan gần như tuyệt đối với vị trí của mục ($r = -0,99$ ở phép cộng), khẳng định trật tự sắp xếp các mục theo độ khó tăng dần là hợp lý.

Bảng 2. Độ khó và độ phân biệt của các mục theo tiểu trắc nghiệm (toàn mẫu)

Tiểu trắc nghiệm	Số mục	Độ khó p (khoảng; TB)	Độ phân biệt rit (khoảng; TB)	Mục rit < 0,20	Mục $p > 0,95$
Phép cộng	50	0,07-1,00 (0,59)	-0,00-0,81 (0,55)	8	10
Phép trừ	50	0,04-0,99 (0,53)	-0,00-0,84 (0,52)	7	8
Phép nhân	50	0,30-1,00 (0,74)	0,07-0,82 (0,50)	11	13
Phép chia	50	0,30-0,99 (0,72)	0,07-0,81 (0,55)	6	9

Ghi chú: p = độ khó (tỉ lệ trả lời đúng); rit = hệ số tương quan mục-tổng hiệu chỉnh; mục chưa kịp thực hiện được tính là chưa đạt. Phép nhân và phép chia: $N = 1.040$ và 1.042 ; phép cộng và phép trừ: $N = 1.342$.

3.2. Hiệu lực phát triển và thống kê mô tả theo cấp lớp

Tiêu chí thứ ba của câu hỏi nghiên cứu thứ nhất là hiệu lực phát triển. Bảng 3 trình bày thống kê mô tả điểm TTA theo từng phép tính và cấp lớp. Nhìn chung, thành tích tăng dần theo cấp lớp, phản ánh quá trình tự động hóa tính toán theo độ tuổi.

Bảng 3. Thống kê mô tả điểm TTA theo cấp lớp và phép tính ($n = 1.342$)

Cấp lớp	Phép tính	Trung bình	Trung vị	Độ lệch chuẩn
Lớp 1	Cộng	12,75	12	3,58
Lớp 1	Trừ	11,15	11	3,32
Lớp 2	Cộng	13,67	12	5,94
Lớp 2	Trừ	12,69	12	4,14
Lớp 3	Cộng	27,15	27	6,74
Lớp 3	Trừ	24,34	25	7,40
Lớp 3	Nhân	30,92	30	8,82
Lớp 3	Chia	28,41	28	11,36
Lớp 4	Cộng	33,53	33	7,26
Lớp 4	Trừ	29,90	30	8,20
Lớp 4	Nhân	39,01	40	9,31
Lớp 4	Chia	37,90	39	10,40
Lớp 5	Cộng	38,01	39	7,74
Lớp 5	Trừ	35,06	36	8,37
Lớp 5	Nhân	42,52	47	8,79
Lớp 5	Chia	41,15	44	10,06
Lớp 6	Cộng	37,96	39	6,77
Lớp 6	Trừ	33,18	34	7,86
Lớp 6	Nhân	35,31	36	11,20
Lớp 6	Chia	35,43	37	12,09

Ghi chú: Theo thiết kế TTA, phép nhân và phép chia không được tiến hành ở lớp 1 và lớp 2; từ lớp 3 thực hiện đủ bốn phép tính.

Đối với phép cộng và phép trừ (lớp 1-6), điểm số tăng mạnh và đều đặn theo cấp lớp; khác biệt giữa các khối có ý nghĩa thống kê với mức tác động rất lớn (cộng: $F(5, 1336) = 555,8$; $\eta^2 = 0,675$; trừ: $F(5, 1336) = 378,0$; $\eta^2 = 0,586$; đều $p < 0,001$) và tương quan thuận chặt chẽ với cấp lớp ($r = 0,78$ và $0,72$) cũng như với tuổi tính theo tháng ($r = 0,76$ và $0,70$). Đối với phép nhân và phép chia (lớp 3-6), điểm số cũng khác biệt có ý nghĩa (nhân: $F(3, 1036) = 61,1$; $\eta^2 = 0,150$; chia: $F(3, 1038) = 54,8$; $\eta^2 = 0,137$; $p < 0,001$) và tăng từ lớp 3 đến lớp 5, song mức tác động nhỏ hơn. Riêng ở hai phép tính này, HS lớp 6 đạt điểm thấp hơn lớp 5 (xem bảng 3); hiện tượng này được bàn luận ở mục 4. Như vậy, cả ba tiêu chí của câu hỏi nghiên cứu thứ nhất đều được đáp ứng, rõ rệt nhất ở phép cộng và phép trừ trong suốt cấp tiểu học; đây là cơ sở để tiến hành xây dựng bộ chuẩn tham chiếu ở mục tiếp theo.

3.3. Chuẩn bách phân vị theo phép tính và cấp lớp

Để trả lời câu hỏi nghiên cứu thứ hai, bảng 4 trình bày chuẩn bách phân vị phục vụ tham chiếu lâm sàng trực tiếp: với một HS ở một cấp lớp nhất định, nhà thực hành đối chiếu điểm TTA của trẻ với hàng tương ứng để xác định vị

trí của trẻ trong nhóm. Chẳng hạn, ở phép cộng lớp 4, ngưỡng Pc16 là 26 điểm; một HS đạt dưới 26 được xem là có thành tích thấp hơn -1 độ lệch chuẩn và thuộc diện cần lưu ý.

Bảng 4. Chuẩn bách phân vị (percentile) điểm TTA theo cấp lớp và phép tính

Cấp lớp	Phép tính	Pc2	Pc 5	Pc 10	Pc 16	Pc 25	Pc 50	Pc 75
Lớp 1	Cộng	6	8	9,10	10	11	12	14
Lớp 1	Trừ	5	7	8	9	9	11	13
Lớp 2	Cộng	6,18	7,55	9	10	10	12	15
Lớp 2	Trừ	7	7,55	9	10	10	12	15
Lớp 3	Cộng	14	17	18	20	23	27	31
Lớp 3	Trừ	10,6	13	16	17	18	25	30
Lớp 3	Nhân	13,5	18	20	23	25	30	36
Lớp 3	Chia	9	11	13,50	16	19	28	37
Lớp 4	Cộng	16,6	22	24	26	29	33	38
Lớp 4	Trừ	13	16	18	20,96	24,75	30	35
Lớp 4	Nhân	18	22	26	29	32	40	48,25
Lớp 4	Chia	14,6	19,55	22	26	31	39	48
Lớp 5	Cộng	20	24	27,40	30	32	39	43
Lớp 5	Trừ	16,7	18,70	23	26,44	30	36	41,50
Lớp 5	Nhân	19,7	25,65	29	32	36	47	50
Lớp 5	Chia	14,6	19,70	25	31	37	44	49
Lớp 6	Cộng	22	27	29	31	34	39	42,50
Lớp 6	Trừ	15	18	22,60	25	28	34	38
Lớp 6	Nhân	13	14	19	23	27,50	36	45
Lớp 6	Chia	7,6	14	18	21	28	37	45

Ghi chú: Pc = bách phân vị. Pc16 là ngưỡng lâm sàng tương đương -1 độ lệch chuẩn theo INAMI; Pc10 và Pc25 là các mốc tham chiếu hỗ trợ. Pc2 = bách phân vị thứ 2, được dùng làm ngưỡng tham chiếu nghiêm ngặt (tương đương khoảng -2 độ lệch chuẩn) để nhận diện trẻ có nguy cơ rối loạn tính toán nặng. Phép nhân và phép chia không áp dụng ở lớp 1 và lớp 2.

Bảng 4 cũng cho thấy, hiệu ứng trần (ceiling effect) ở các phép tính dễ và ở các bách phân vị cao: tại nhiều cấp lớp, giá trị Pc75 tiệm cận hoặc đạt mức tối đa 50 điểm (ví dụ phép nhân, phép chia ở lớp 4 và lớp 5). Ý nghĩa của đặc điểm này đối với việc sử dụng bộ chuẩn được bàn luận ở mục 4.

4. Kết luận và bình luận

Bài báo đã thích nghi trắc nghiệm sàng lọc TTA sang bối cảnh Việt Nam và hiệu chuẩn trên 1342 HS lớp 1-6 tại TP. Hồ Chí Minh. Đối với câu hỏi thứ nhất, cả 03 tiêu chí tâm trắc đều đạt: độ tin cậy nhất quán nội tại rất cao ở từng tiêu trắc nghiệm và toàn thang, các mục có độ khó trải rộng và độ phân biệt tốt, thành tích tăng theo cấp lớp với mức tác động lớn ở phép cộng và phép trừ. Đối với câu hỏi thứ hai, tác giả đã thiết lập bộ chuẩn tham chiếu bách phân vị đầu tiên cho TTA phiên bản Việt Nam theo từng phép tính và cấp lớp, kèm hướng dẫn vận dụng các mốc Pc16, Pc10, Pc25 và Pc2. Đóng góp khoa học của bài báo nằm ở 03 điểm, cụ thể: *Thứ nhất*, đây là bộ chuẩn tham chiếu định lượng đầu tiên cho một công cụ lượng giá xử lý số học được xây dựng trên dữ liệu trẻ em tại TP. Hồ Chí Minh, Việt Nam, qua đó chuyển việc nhận diện HS có khó khăn tính toán từ nhận định chủ quan của GV sang một cơ sở đối chiếu định lượng; *Thứ hai*, bài báo cung cấp bằng chứng cho thấy một công cụ đo lường tính toán phát triển ở châu Âu vẫn giữ được đặc tính tâm trắc tốt trong bối cảnh giáo dục và ngôn ngữ Việt Nam, củng cố giá trị của cấu trúc lưu loát tính toán như một chỉ báo sàng lọc xuyên văn hóa; *Thứ ba*, bài báo đề xuất một mô hình phân tầng ngưỡng phù hợp với nguồn lực của trường học tại Việt Nam thay vì áp dụng máy móc một điểm cắt duy nhất.

Đặt trong tương quan quốc tế, các đặc tính tâm trắc của phiên bản Việt Nam nằm trong khoảng đã được báo cáo với những công cụ tương đương. Karagiannakis và Noël (2020) khi đánh giá bằng trắc nghiệm MathPro trên 622 HS lớp 1-6 tại Bỉ cũng ghi nhận điểm số tăng đều theo cấp lớp và độ nhất quán nội tại thỏa đáng - mô thức trùng khớp với mẫu Việt Nam. Gliksman và cộng sự (2022), khi hiệu chuẩn một trắc nghiệm lưu loát tính toán trên HS lớp 1-6 tại Israel, cho thấy tính lưu loát tiếp tục phát triển suốt cấp tiểu học, đồng thời mức thành thực khác nhau rõ rệt giữa bốn phép tính và tính tự động hóa của mỗi phép tính được xác lập ở những cấp lớp khác nhau. Phát hiện này lý giải trực tiếp vì sao trong mẫu Việt Nam, mức tác động của cấp lớp lên phép cộng và phép trừ ($\eta^2 = 0,675$ và $0,586$) lớn hơn hẳn so với phép nhân và phép chia ($\eta^2 = 0,150$ và $0,137$): cộng và trừ được học sớm và luyện tập liên tục qua sáu năm nên quỹ đạo tự động hóa trải dài, trong khi nhân và chia chỉ bắt đầu từ lớp 3, khoảng quan sát ngắn hơn và

điểm số nhanh chóng tiệm cận trần. Kết quả của Lafay và cộng sự (2015) tại Québec cho thấy, trẻ Québec đạt điểm thấp hơn trẻ Hà Lan trên cùng một công cụ, điều này củng cố luận điểm cốt lõi của bài báo này: chuẩn tham chiếu không thể vay mượn giữa các bối cảnh mà bắt buộc phải được xây dựng cho từng quốc gia.

02 đặc điểm của bộ dữ liệu cần lưu ý khi sử dụng bộ chuẩn là: *Thứ nhất* là hiệu ứng trần ở các phép tính dễ và vùng bách phân vị cao, khiến công cụ kém nhạy trong phân hóa nhóm HS khá - giỏi. Đây là hệ quả tất yếu của thiết kế: một trắc nghiệm sàng lọc được tối ưu cho vùng điểm thấp, nơi quyết định lâm sàng được đưa ra, chứ không nhằm xếp hạng HS giỏi; *Thứ hai* là hiện tượng HS lớp 6 đạt điểm phép nhân và phép chia thấp hơn lớp 5. Cách lý giải khả dĩ nhất mang tính phương pháp: HS lớp 6 trong mẫu được lấy từ các trường THCS khác với các trường tiểu học cung cấp dữ liệu lớp 1-5, nên khác biệt có thể phản ánh khác biệt giữa các trường hơn là một quy luật phát triển; do thiết kế cắt ngang, khả năng này không thể loại trừ bằng dữ liệu hiện có. Bên cạnh đó, bằng chứng thực nghiệm cho thấy quỹ đạo phát triển của tính lưu loát không đơn điệu và khác nhau giữa các phép tính: Gliksman và cộng sự (2022) ghi nhận tính tự động hóa của mỗi phép tính đạt tới ở những cấp lớp khác nhau và tốc độ cải thiện chậm lại sau giai đoạn luyện tập tập trung; Thevenot và cộng sự (2023) chỉ ra năng lực ở một phép tính phụ thuộc vào các phép tính đã học trước đó và mỗi phụ thuộc này thay đổi theo cấp lớp - phép cộng dự báo phép chia ở lớp 4 nhưng không còn ở lớp 5. Khi lên lớp 6, trọng tâm môn Toán chuyển sang số nguyên, phân số, biểu thức đại số và hình học nên các phép tính cơ bản ít được luyện tập trực tiếp hơn. Đây là các giả thuyết cạnh tranh chưa được kiểm định trên dữ liệu hiện có; hệ quả thực hành là chuẩn của lớp 6 nên được sử dụng thận trọng hơn chuẩn của các lớp tiểu học.

Về cách vận dụng bộ chuẩn, mốc Pc16 (tương đương -1 độ lệch chuẩn) theo quy định của INAMI nên được dùng làm ngưỡng sàng lọc chính, bởi việc ưu tiên ngưỡng -1 độ lệch chuẩn trong lĩnh vực nhận thức số giúp tối ưu hóa độ nhạy, bảo đảm không bỏ sót những HS mới bắt đầu bộc lộ khó khăn trong tiến trình tự động hóa. Để tăng hiệu quả phân loại trong thực tế học đường, nhà thực hành có thể kết hợp đối chiếu với mốc Pc2 (xấp xỉ -2 độ lệch chuẩn), tạo ra một cơ chế phân tầng nguy cơ rõ rệt: nếu Pc16 - cùng các mốc hỗ trợ Pc25 và Pc10 - đóng vai trò khoanh vùng nhóm HS “có nguy cơ” cần được GV chú ý hỗ trợ sự phạm hoặc phụ đạo giảm tải tại lớp, thì Pc2 là chỉ báo cho các trường hợp khiếm khuyết sâu sắc về lưu loát tính toán, đòi hỏi kế hoạch can thiệp chuyên sâu và phối hợp liên ngành. Dù vận dụng điểm cắt nào tùy theo nguồn lực của từng cơ sở giáo dục, cần nhấn mạnh rằng TTA chỉ giữ vai trò một trắc nghiệm sàng lọc nhanh: một kết quả dưới ngưỡng chỉ là tín hiệu cảnh báo nguy cơ, là cơ sở để nhà trường chuyển HS sang bước lượng giá chẩn đoán chuyên sâu đa lĩnh vực, chứ không thể dùng làm bằng chứng độc lập để đưa ra một nhãn chẩn đoán rối loạn tính toán chính thức (Lafay và cộng sự, 2014; Noël, 2005).

Một trường hợp đã ẩn danh minh họa vị trí của TTA trong thực hành. Em N. (nữ, 8 tuổi, lớp 3) được cha mẹ đưa đến đánh giá với lời than phiền “toán chậm”; trẻ có trí tuệ bình thường (WISC-IV) và tiền sử chậm phát triển ngôn ngữ sớm đã được can thiệp. Đối chiếu với chuẩn lớp 3 trong bảng 4, kết quả TTA cho thấy thành tích rất thấp ở cả bốn phép tính (cộng 11/50, trừ 13/50, nhân 12/50, chia 5/50), đều rơi vào dưới Pc2 hoặc dưới Pc5. Quan sát định tính bổ sung cho thấy trẻ phải kết hợp đếm ngón tay và lẫn lộn phép cộng với phép trừ (ví dụ $20 - 8 = 28$). Như vậy, TTA đã hoàn thành đúng chức năng sàng lọc nguy cơ và định vị mức độ nghiêm trọng; bước tiếp theo cần sử dụng trắc nghiệm chẩn đoán chuyên sâu (Van Nieuwenhoven và cộng sự, 2002) để xác định cơ chế của khó khăn.

Từ kết quả trên có thể rút ra một số khuyến nghị cho thực tiễn giáo dục và tâm lý học đường. TTA cùng bộ chuẩn này phù hợp nhất khi được triển khai như một bước sàng lọc phổ quát định kỳ ở cấp lớp học: GV hoặc chuyên viên tâm lý học đường thực hiện cho cả lớp trong khoảng 10 phút vào đầu hoặc giữa năm học, rồi đối chiếu điểm từng HS với hàng tương ứng trong bảng 4; HS dưới Pc16 đưa vào diện theo dõi và hỗ trợ tăng cường tại lớp, HS dưới Pc2 chuyển tới lượng giá chuyên sâu. Kết quả nên được đọc theo từng phép tính chứ không gộp thành một điểm tổng, vì hồ sơ khó khăn có thể chỉ khu trú ở một hoặc hai phép tính. Về phạm vi áp dụng, cần lưu ý 03 giới hạn: bộ chuẩn được xây dựng trên HS tại TP. Hồ Chí Minh nên chỉ áp dụng cho bối cảnh có điều kiện dạy học tương đồng, dùng cho vùng miền khác cần hết sức thận trọng cho tới khi có dữ liệu đa địa bàn; chuẩn lập theo cấp lớp chứ không theo tháng tuổi nên với HS học sớm hoặc học muộn cần đối chiếu thêm chuẩn của lớp liền kề; và mọi mốc bách phân vị ở đây là ngưỡng sàng lọc, không phải tiêu chí chẩn đoán.

Kết quả bài báo cần được diễn giải trong phạm vi những giới hạn của nghiên cứu. Mẫu được thu thập tại một địa bàn duy nhất theo phương pháp chọn mẫu thuận tiện theo cụm nên tính đại diện cho HS cả nước còn hạn chế; bài báo mới dừng ở chuẩn tham chiếu mô tả, chưa có nhóm lâm sàng đã được chẩn đoán làm đối chứng để xác định điểm cắt tối ưu (độ nhạy - độ đặc hiệu, đường cong ROC); tỉ lệ thiếu thông tin giới tính (14%) hạn chế phân tích khác biệt theo giới; và thiết kế cắt ngang không cho phép loại trừ hiệu ứng trường học khi diễn giải khác biệt giữa lớp 6 và lớp 5. Dù dữ liệu được thu thập từ năm 2018, năng lực tự động hóa các phép tính cơ bản là tiến trình nhận thức

nền tảng, ít biến động theo thay đổi của chương trình học, nên bộ dữ liệu vẫn giữ giá trị tham chiếu ban đầu. Các nghiên cứu tiếp theo cần mở rộng mẫu đa địa bàn có phân tầng, kiểm định độ hiệu lực hội tụ với các trắc nghiệm chẩn đoán chuyên sâu, thu thập dữ liệu trên nhóm lâm sàng để xác lập điểm cắt tối ưu, và theo dõi dọc qua giai đoạn chuyển cấp nhằm làm sáng tỏ hiện tượng chững lại của tính lưu loát ở lớp 6.

Tuyên bố về GenAI và Quyền tác giả: Trong quá trình chuẩn bị bản thảo, tác giả có sử dụng công cụ trí tuệ nhân tạo Claude 3 Opus nhằm hỗ trợ rà soát diễn đạt, tránh lặp từ và cải thiện tính mạch lạc của câu văn.

Tuyên bố về xung đột lợi ích: Tác giả tuyên bố không có xung đột lợi ích.

Lời cảm ơn: Nghiên cứu nằm trong khuôn khổ dự án “Nhận diện, lượng giá và can thiệp rối loạn chuyên biệt học tập trong bối cảnh học đường tại Việt Nam” do Trường Đại học Khoa học Xã hội và Nhân văn (Đại học Quốc gia Thành phố Hồ Chí Minh), UCLouvain và ULiège (Vương quốc Bỉ) phối hợp thực hiện. Tác giả cảm ơn các học viên của dự án đào tạo đã hỗ trợ thu thập dữ liệu, cùng các nhà trường, thầy cô và học sinh đã tham gia.

Tài liệu tham khảo

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Cohen Kadosh, R., Dowker, A., Heine, A., & Kaufmann, L. (2013). Interventions for improving numerical abilities: Present and future. *Trends in Neuroscience and Education*, 2(2), 85-93. <https://doi.org/10.1016/j.tine.2013.04.001>
- Cook, B. G., Tankersley, M., & Landrum, T. J. (2013). Evidence-based practices in learning and behavioral disabilities: The search for effective instruction. In B. G. Cook, M. Tankersley, & T. J. Landrum (Eds.), *Evidence-based practices* (Advances in Learning and Behavioral Disabilities, Vol. 26, pp. 1-19). Emerald Group Publishing.
- de Vos, T. (2010). *TempoTest Automatiseren (TTA)*. Boom Test Uitgevers.
- Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, 37(1), 4-15. <https://doi.org/10.1177/00222194040370010201>
- Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities*, 45(4), 291-307. <https://doi.org/10.1177/0022219410392046>
- Gliksman, Y., Berebbi, S., & Henik, A. (2022). Math fluency during primary school. *Brain Sciences*, 12(3), Article 371. <https://doi.org/10.3390/brainsci12030371>
- Institut National d'assurance maladie-invalidité (2024). *Nomenclature des prestations de santé: Chapitre X - Logopédie, Article 36*. <https://www.inami.fgov.be>
- Karagiannakis, G., & Noël, M.-P. (2020). Mathematical Profile Test: A preliminary evaluation of an online assessment for mathematics skills of children in Grades 1-6. *Behavioral Sciences*, 10(8), 126. <https://doi.org/10.3390/bs10080126>
- Lafay, A., Saint-Pierre, M.-C., & Macoir, J. (2014). L'évaluation des habiletés mathématiques de l'enfant: Inventaire critique des outils disponibles. *Glossa*, 116, 33-58.
- Lafay, A., St-Pierre, M.-C., & Macoir, J. (2015). Validation franco-qubécoise du Tempo Test Rekenen pour l'évaluation des habiletés mathématiques auprès d'enfants de 8-9 ans. *Glossa*, 118, 27-39.
- Lê Thị Mai Liên (2018). *Trắc nghiệm sàng lọc và chẩn đoán rối loạn tính toán*. Kỷ yếu Hội thảo Khoa học quốc tế Tâm lý học đường (tr 106-118), Trường Đại học Sư phạm Hà Nội.
- Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, 40(5), 458-478. <https://doi.org/10.1177/00222194070400050901>
- Noël, M.-P. (2005). *La dyscalculie: Trouble du développement numérique de l'enfant*. Solal.
- Thevenot, C., Tazouti, Y., Billard, C., Dewi, J., & Fayol, M. (2023). Acquisition of new arithmetic skills based on prior arithmetic skills: A cross-sectional study in primary school from grade 2 to grade 5. *British Journal of Educational Psychology*, 93(3), 727-741. <https://doi.org/10.1111/bjep.12588>
- Van Nieuwenhoven, C., Grégoire, J., & Noël, M.-P. (2002). *TEDI-MATH: Test diagnostique des compétences de base en mathématiques*. ECPA.
- Wang, Y., Long, J., & Wang, P. (2024). The prevalence of mathematical difficulties among primary school children in Mainland China: A systematic review and meta-analysis. *Frontiers in Public Health*, 11, 1250337. <https://doi.org/10.3389/fpubh.2023.1250337>