VJE Tạp chí Giáo dục (2026), 26(20), 8-14 ISSN: 2354-0753

ĐÁNH GIÁ ĐA NGUỒN NĂNG LỰC GIAO TIẾP VÀ HỢP TÁC TRONG MÔI TRƯỜNG SỐ CỦA HỌC SINH TIỂU HỌC: XU HƯỚNG THAY ĐỔI CỦA KẾT QUẢ ĐÁNH GIÁ QUA BỐN BÀI HỌC CAN THIỆP

MULTI-SOURCE ASSESSMENT OF DIGITAL COMMUNICATION AND COLLABORATION COMPETENCE AMONG PRIMARY SCHOOL STUDENTS: TRENDS IN ASSESSMENT RESULTS ACROSS FOUR INTERVENTION LESSONS

Cao Hồng Huệ[1,+], Đào Thái Lai[2], Nguyễn Thị Thanh Huyền[3]

[1]Trường Đại học Sư phạm Hà Nội 2;
[2]Hội Khoa học Tâm lí - Giáo dục Việt Nam;
[3]Trường Đại học Sư phạm Hà Nội
+*Tác giả liên hệ ● Email: caohonghue@hpu2.edu.vn*

Article history
Received: 05/8/2026
Accepted: 10/9/2026
Published: 20/10/2026

Keywords
Multi-source assessment, digital communication and collaboration competence, peer assessment, self-assessment, digital competence

ABSTRACT
Digital communication and collaboration competence (DCCC) is a core component of digital competence among primary school students. Self-assessment, the most commonly used approach for assessing this competence, has limited reliability and remains insufficiently validated for primary school students. Multi-source assessment has been considered a feasible alternative; however, empirical evidence remains limited, particularly studies that concurrently examine multiple assessment sources over time. This study describes and compares trends in DCCC assessment results among primary school students using three assessment sources: student self-assessment, peer assessment, and teacher assessment, during a pedagogical intervention in primary school ICT lessons. A quasi-experimental design was employed across eight primary schools, with the present analysis focusing on descriptive data from the intervention group. Data were collected repeatedly after four consecutive intervention lessons and analyzed using descriptive statistics. The results showed that all three assessment sources recorded a broadly consistent upward trend, with the most evident change occurring at the same stage across the three sources. At the same time, differences in absolute scores persisted among the assessment sources, being most pronounced during the early stages of the intervention, decreasing sharply after the first lesson before fluctuating slightly in the final two lessons. These findings indicate that combining multiple assessment sources is a feasible approach, providing additional empirical evidence for monitoring trends in DCCC assessment results among students in primary school ICT lessons. They also suggest that the results of multi-source assessment should be interpreted cautiously during the early stages of an instructional intervention.

1. Mở đầu

Nghiên cứu về năng lực giao tiếp và hợp tác trong môi trường số (GT&HTS) đã gia tăng nhanh chóng trong những năm gần đây, phản ánh vai trò ngày càng quan trọng của thành phần năng lực này trong cấu trúc năng lực số (Cao và cộng sự, 2025). Trong khung năng lực số dành cho HS tiểu học, năng lực GT&HTS được xác định là một thành phần cốt lõi, gồm sáu thành phần năng lực phản ánh các khía cạnh khác nhau của giao tiếp, hợp tác và tham gia trong môi trường số (Cao Hồng Huệ và cộng sự, 2025). Quan niệm này cũng phù hợp với các nghiên cứu quốc tế, trong đó năng lực giao tiếp và hợp tác được xem là một thành phần của năng lực số ở HS tiểu học (León-Sandobal và cộng sự, 2025). Sự quan tâm đối với năng lực số không chỉ diễn ra ở phạm vi quốc tế mà còn ngày càng rõ nét tại Việt Nam. Một số nghiên cứu gần đây đã vận dụng khung DigComp để đánh giá năng lực số của công dân Việt Nam và xác nhận tính phù hợp của khung này trong bối cảnh Việt Nam (Nguyen và cộng sự, 2024). Tuy nhiên, mức độ phát triển về lí luận vẫn chưa tương xứng với các nghiên cứu thực nghiệm, bởi số lượng nghiên cứu đánh giá năng

lực số ở HS tiểu học còn tương đối hạn chế (Godaert và cộng sự, 2022). Khoảng cách này đặt ra yêu cầu xem xét lại công cụ đang được sử dụng phổ biến nhất để đánh giá năng lực ở lứa tuổi này, đó là tự đánh giá.

Tự đánh giá là công cụ được sử dụng phổ biến nhất để đánh giá năng lực GT&HTS ở HS tiểu học nhờ tính đơn giản trong triển khai và sự sẵn có của khung năng lực làm cơ sở xây dựng công cụ (Cao Hồng Huệ và cộng sự, 2025). Tuy nhiên, ưu điểm về tính khả thi này cũng đi kèm một hạn chế đã được ghi nhận khá nhất quán trong các nghiên cứu trước: HS thường có xu hướng đánh giá năng lực số của bản thân cao hơn so với biểu hiện thực tế (Porat và cộng sự, 2018) và hiện tượng thiên lệch tương tự cũng được ghi nhận ở người học nói chung khi tự đánh giá kĩ năng số (Černochová và cộng sự, 2020). Điều đáng lưu ý là các nghiên cứu kiểm chứng độ tin cậy của tự đánh giá ở đúng đối tượng HS tiểu học vẫn còn tương đối ít (Godaert và cộng sự, 2022). Trên cơ sở đó, một nghiên cứu trước của nhóm tác giả đã đối chiếu kết quả tự đánh giá của HS với đánh giá của GV trên một mẫu quy mô lớn và ghi nhận khoảng lệch đáng kể giữa hai nguồn tại một thời điểm cắt ngang (Cao Hồng Huệ và cộng sự, 2026). Nghiên cứu hiện tại kế thừa công cụ tự đánh giá và rubric đánh giá của GV đã được xây dựng và kiểm định trong nghiên cứu trước đó, nhưng có hai điểm khác biệt cơ bản: (1) Thay vì sử dụng dữ liệu cắt ngang tại một thời điểm, nghiên cứu khai thác dữ liệu được thu thập lặp lại ở bốn thời điểm liên tiếp trong một thực nghiệm sư phạm cụ thể, nhằm mô tả xu hướng thay đổi của kết quả đánh giá theo thời gian; (2) Thay vì chỉ so sánh hai nguồn đánh giá, nghiên cứu bổ sung thêm đánh giá đồng đẳng để đối chiếu đồng thời ba nguồn, qua đó mở rộng câu hỏi nghiên cứu sang vai trò của đánh giá đa nguồn trong việc khắc phục hạn chế của đánh giá đơn nguồn.

Đánh giá đa nguồn, kết hợp nhận định từ nhiều chủ thể như bản thân HS, bạn học và GV, được xem là một hướng tiếp cận có tiềm năng khắc phục những hạn chế của đánh giá đơn nguồn. Xu hướng tăng cường sự tham gia của nhiều chủ thể trong đánh giá cũng được ghi nhận trong các tổng quan hệ thống về thực hành đánh giá hợp tác trong giáo dục (Widodo và cộng sự, 2025). Cùng với đó, các phương pháp đánh giá đổi mới, như đánh giá đồng đẳng và tự đánh giá, ngày càng được quan tâm và vận dụng trong thực tiễn giáo dục (Devaki, 2024). Tuy nhiên, sự phát triển về mặt lí luận này vẫn chưa đi kèm với các bằng chứng thực nghiệm tương xứng. Vì vậy, mặc dù được khuyến nghị rộng rãi về mặt lí luận, đánh giá đa nguồn đối với năng lực GT&HTS của HS tiểu học vẫn thiếu những bằng chứng thực nghiệm, đặc biệt trong các nghiên cứu theo dõi đồng thời nhiều nguồn đánh giá độc lập trong cùng một bối cảnh dạy học.

Phần lớn các nghiên cứu hiện có, bao gồm cả nghiên cứu trước của nhóm tác giả, chỉ xem xét năng lực GT&HTS tại một thời điểm duy nhất. Do đó, bên cạnh khoảng trống về đánh giá đa nguồn còn tồn tại một khoảng trống khác là thiếu các nghiên cứu theo dõi sự thay đổi trong kết quả đánh giá năng lực này qua nhiều thời điểm trong cùng một quá trình dạy học. Tuy nhiên, thiết kế nghiên cứu với nhiều lần đo trên cùng một nhóm HS đã được sử dụng để theo dõi sự phát triển kĩ năng ở bậc tiểu học trong các lĩnh vực khác (Claes và cộng sự, 2026), cho thấy cách tiếp cận này khả thi nhưng chưa được vận dụng để nghiên cứu sự thay đổi của kết quả đánh giá năng lực GT&HTS theo thời gian. Năng lực GT&HTS của nhóm thực nghiệm được ghi nhận đồng thời từ ba nguồn đánh giá độc lập gồm tự đánh giá, đánh giá đồng đẳng và đánh giá của GV tại bốn thời điểm liên tiếp, tương ứng với bốn bài học can thiệp. Cách tổ chức thu thập dữ liệu này tạo điều kiện để đối chiếu ba nguồn đánh giá theo thời gian, qua đó góp phần lấp đầy cả hai khoảng trống nghiên cứu đã nêu.

Trên cơ sở đó, nghiên cứu này nhằm mô tả và đối chiếu xu hướng thay đổi trong kết quả đánh giá năng lực GT&HTS của HS tiểu học thông qua ba nguồn đánh giá độc lập trong quá trình thực nghiệm sư phạm. Cụ thể, nghiên cứu tập trung trả lời hai câu hỏi: (1) Kết quả đánh giá năng lực GT&HTS được ghi nhận từ ba nguồn đánh giá qua bốn bài học can thiệp có đặc điểm gì về xu hướng và thời điểm chuyển biến? (2) Ba nguồn đánh giá có những điểm giống và khác nhau như thế nào trong xu hướng thay đổi của kết quả đánh giá, và điều này có ý nghĩa gì đối với việc vận dụng phối hợp đánh giá đa nguồn trong dạy học Tin học ở tiểu học? Nghiên cứu đóng góp trên ba phương diện: (1) Cung cấp bằng chứng thực nghiệm về việc theo dõi đồng thời ba nguồn đánh giá theo thời gian - hướng tiếp cận mà các nghiên cứu trước, kể cả nghiên cứu trước của nhóm tác giả, chưa thực hiện; (2) Làm rõ những điểm giống và khác nhau trong xu hướng thay đổi của kết quả đánh giá giữa các nguồn, qua đó cung cấp cơ sở thực nghiệm để xem xét việc phối hợp nhiều nguồn đánh giá trong đánh giá năng lực số; (3) Cung cấp cơ sở thực tiễn để GV Tin học tiểu học tham khảo khi vận dụng phối hợp các nguồn đánh giá trong quá trình dạy học.

2. Phương pháp nghiên cứu

Bài báo này trình bày một phân tích mô tả dựa trên dữ liệu của nhóm thực nghiệm, được thu thập trong khuôn khổ một nghiên cứu bán thực nghiệm (quasi-experimental) rộng hơn, sử dụng thiết kế trước - sau có nhóm đối chứng nhằm đánh giá sự phát triển của năng lực GT&HTS của HS trong bối cảnh dạy học thực tế. Nhóm đối chứng thuộc

thiết kế thực nghiệm tổng thể nhưng không được phân tích trong bài báo này. Tại mỗi trường tham gia, một lớp được chọn làm nhóm thực nghiệm (EXP), triển khai các biện pháp can thiệp sư phạm trong bốn bài học liên tiếp; một lớp khác làm nhóm đối chứng, duy trì cách tổ chức dạy học hiện hành trong cùng khoảng thời gian. Do đặc điểm tổ chức dạy học ở cấp tiểu học, việc phân nhóm được thực hiện theo các lớp học đã hình thành từ đầu năm học, không phân ngẫu nhiên từng HS. Thực nghiệm được triển khai tại 8 trường tiểu học thuộc 5 tỉnh, thành phố trong học kì II năm học 2025-2026.

Đối tượng nghiên cứu là HS của 8 lớp thực nghiệm thuộc ba khối lớp 3, 4 và 5 tại 8 trường tham gia nghiên cứu. Dữ liệu được thu thập lặp lại sau mỗi bài học can thiệp thông qua ba nguồn đánh giá độc lập, vì vậy đơn vị phân tích không đồng nhất giữa các nguồn: tự đánh giá (SA) và đánh giá đồng đẳng (PA) được ghi nhận ở cấp độ cá nhân HS, trong khi đánh giá của GV (TR) được ghi nhận ở cấp độ nhóm học tập. Cỡ mẫu của từng nguồn giảm dần qua các bài học, từ 308 xuống còn 74 lượt đánh giá đối với SA, từ 305 xuống còn 71 lượt đối với PA và từ 27 xuống còn 5 lượt đối với TR. Sự suy giảm này phản ánh số trường và số lớp thực nghiệm tham gia đầy đủ ở từng bài học không hoàn toàn giống nhau: 8 trường/8 lớp ở Bài học 1, 6 trường/6 lớp ở Bài học 2, 5 trường/5 lớp ở Bài học 3 và 2 trường/2 lớp ở Bài học 4, do điều kiện triển khai thực tế tại các trường tham gia khiến một số trường không thể tiếp tục triển khai toàn bộ chuỗi bài học thực nghiệm. Riêng đối với TR, sau mỗi bài học, GV lựa chọn ngẫu nhiên khoảng 2-4 nhóm HS để thực hiện đánh giá, ưu tiên luân phiên giữa các nhóm khi điều kiện cho phép nhằm tạo cơ hội để mỗi nhóm được GV đánh giá ít nhất một lần trong quá trình thực nghiệm. Do đó, nghiên cứu không theo dõi cố định cùng một nhóm HS qua cả bốn bài học và dữ liệu TR nên được hiểu là các quan sát độc lập tại từng thời điểm hơn là một chuỗi theo dõi lặp lại trên cùng một nhóm cố định. Đặc điểm mẫu và sự biến động cỡ mẫu của từng nguồn được trình bày tại bảng 1.

Ba công cụ đánh giá được sử dụng để thu thập dữ liệu về năng lực GT&HTS của HS, được ghi nhận trong quá trình dạy học môn Tin học. Thang tự đánh giá (SA) và rubric đánh giá của GV (TR) đều gồm 18 biểu hiện, phản ánh đầy đủ sáu thành phần của khung năng lực GT&HTS dành cho HS tiểu học. Hai công cụ này được xây dựng và kiểm định trong một nghiên cứu trước của nhóm tác giả (Cao Hồng Huệ và cộng sự, 2026), do đó bài báo hiện tại không trình bày lại quy trình phát triển và kiểm định thang đo. Bên cạnh đó, nghiên cứu sử dụng phiếu đánh giá đồng đẳng (PA) do nhóm nghiên cứu thiết kế riêng cho thực nghiệm sư phạm này, gồm bốn tiêu chí để HS đánh giá lẫn nhau sau mỗi bài học. Bốn tiêu chí của PA gồm: chia sẻ ý tưởng, hợp tác làm việc, giao tiếp lịch sự và hoàn thành nhiệm vụ. Mỗi tiêu chí được HS đánh giá theo hình thức số sao (1-4 sao), quy đổi tương ứng thành thang điểm 0-3 (1 sao = 0 điểm, 2 sao = 1 điểm, 3 sao = 2 điểm, 4 sao = 3 điểm). Trước khi thực hiện PA, GV hướng dẫn HS cách sử dụng phiếu trong buổi thực nghiệm đầu tiên; phiếu được thiết kế đơn giản, chủ yếu ở dạng tích chọn, phù hợp với HS tiểu học. Mỗi HS được tất cả các thành viên còn lại trong nhóm đánh giá (không tự đánh giá bản thân), và điểm PA cuối cùng của mỗi HS được tính bằng điểm trung bình cộng của các đánh giá nhận được từ các thành viên khác trong nhóm. Phiếu PA được xây dựng dựa trên khung năng lực GT&HTS của nghiên cứu, nhưng chưa trải qua quy trình thẩm định chuyên gia độc lập hoặc nghiên cứu pilot trước khi áp dụng; hạn chế này được thảo luận thêm ở mục 4. Không giống hai công cụ còn lại, PA chỉ đánh giá ba thành phần của khung năng lực gồm SHR (Chia sẻ thông tin và nội dung thông qua công nghệ số), COL (Hợp tác thông qua công nghệ số) và NET (Quy tắc ứng xử trên môi trường số), chưa bao phủ ba thành phần còn lại là INT (Tương tác thông qua công nghệ số), ENG (Tham gia và đóng góp với tư cách công dân thông qua công nghệ số) và MDI (Quản lí sự hiện diện và danh tính số). Sự khác biệt về phạm vi đo lường này được xem là một giới hạn cần lưu ý khi đối chiếu kết quả giữa ba nguồn đánh giá trong phần thảo luận. Mặt khác, mặc dù cả ba công cụ đều sử dụng chung thang điểm quy đổi 0-3, sự thống nhất về hình thức thang đo này không đồng nghĩa với sự tương đương về cấu trúc đo lường giữa SA, PA và TR. Do đó, chênh lệch điểm số giữa các nguồn cần được hiểu là phản ánh sự khác biệt về góc nhìn và phạm vi đo, không phải mức độ chính xác hoặc mức độ nghiêm khắc khác nhau giữa các nguồn khi đánh giá cùng một năng lực.

Thực nghiệm được triển khai ở ba khối lớp 3, 4 và 5; mỗi khối thực hiện bốn bài học thực nghiệm liên tiếp, tương ứng với bốn thời điểm thu thập dữ liệu. Do nội dung chuyên môn của chương trình Tin học khác nhau giữa các khối lớp, bài báo không trình bày kết quả theo tên từng bài học cụ thể mà chuẩn hóa thành bốn mốc Bài học 1-4, trong đó mỗi mốc tương ứng với bài thực nghiệm thứ nhất, thứ hai, thứ ba và thứ tư của từng khối. Do đó, Bài học 1-4 là bốn thời điểm can thiệp được chuẩn hóa giữa ba khối lớp, không phải bốn bài học duy nhất của toàn bộ nghiên cứu. Mỗi bài học kéo dài hai tiết học (khoảng 70 phút). Mặc dù khác nhau về nội dung chuyên môn, các bài học đều được thiết kế theo cùng một định hướng phát triển năng lực GT&HTS, trong đó HS làm việc theo nhóm, trao đổi và phân công nhiệm vụ, xây dựng và chia sẻ sản phẩm số, nhận xét và phản hồi lẫn nhau; ở một số bài học có tích hợp AI, HS

được tiếp cận các phản hồi do AI tạo ra (hoặc trực tiếp sử dụng AI dưới sự hướng dẫn của GV khi điều kiện cho phép) để tham khảo, so sánh và kiểm chứng thông tin theo quy trình 5P trước khi hoàn thiện sản phẩm. Sau mỗi bài học, HS thực hiện tự đánh giá và đánh giá đồng đẳng, đồng thời GV lựa chọn một số nhóm để thực hiện đánh giá theo quy trình được mô tả ở trên.

Trong quá trình triển khai bốn bài học can thiệp, dữ liệu của nhóm thực nghiệm được thu thập lặp lại sau mỗi bài học thông qua ba nguồn đánh giá: tự đánh giá, đánh giá đồng đẳng và đánh giá của GV. Cả ba nguồn đều được thực hiện ngay sau khi kết thúc từng bài học nhằm bảo đảm các kết quả đánh giá phản ánh cùng một thời điểm đo lường đối với HS. Việc thu thập đồng thời ba nguồn đánh giá giúp hạn chế ảnh hưởng của yếu tố thời gian giữa các phép đo, đồng thời tạo điều kiện đối chiếu kết quả từ các góc nhìn khác nhau đối với cùng một hoạt động học tập. Quy trình này được lặp lại thống nhất qua cả bốn bài học, hình thành bốn thời điểm thu thập dữ liệu liên tiếp trong toàn bộ quá trình thực nghiệm.

Dữ liệu được xử lí bằng thống kê mô tả nhằm phản ánh xu hướng thay đổi của kết quả đánh giá năng lực GT&HTS trong dạy học môn Tin học qua từng bài học can thiệp. Đối với mỗi nguồn đánh giá, nghiên cứu tính toán số lượng quan sát (n), giá trị trung bình (M) và độ lệch chuẩn (SD) theo từng thời điểm thu thập dữ liệu. Xu hướng kết quả đánh giá của ba nguồn được trực quan hóa bằng biểu đồ đường để hỗ trợ so sánh sự thay đổi của điểm trung bình qua bốn bài học. Do mục tiêu của bài báo là mô tả và đối chiếu xu hướng thay đổi của kết quả đánh giá năng lực giữa ba nguồn đánh giá trong bối cảnh thực nghiệm sư phạm, nghiên cứu không thực hiện các kiểm định suy luận thống kê hoặc tính toán độ lớn hiệu ứng đối với các dữ liệu này. Do số quan sát của từng nguồn đánh giá không đồng đều giữa các bài học, phân tích được thực hiện trên các trường hợp có dữ liệu hợp lệ tại mỗi thời điểm; nghiên cứu không thực hiện nội suy hoặc bù dữ liệu thiếu.

3. Kết quả nghiên cứu

3.1. Đặc điểm dữ liệu theo ba nguồn đánh giá

Bảng 1. Số lượng quan sát theo nguồn đánh giá và bài học can thiệp (nhóm EXP)

Nguồn đánh giá	Đơn vị phân tích	Bài học 1	Bài học 2	Bài học 3	Bài học 4
SA (Tự đánh giá)	Cá nhân HS	308	207	165	74
PA (Đánh giá đồng đẳng)	Cá nhân HS	305	166	158	71
TR (Đánh giá của GV)	Nhóm học tập	27	22	19	5

Bảng 1 trình bày số lượng quan sát của ba nguồn đánh giá qua bốn bài học can thiệp. Số lượng quan sát của cả ba nguồn giảm dần qua các bài học do số lượng trường và lớp tham gia thực nghiệm ở từng bài học không hoàn toàn giống nhau; có trường triển khai đầy đủ bốn bài học, trong khi có trường chỉ tham gia một phần của quá trình thực nghiệm. Đối với nguồn đánh giá của GV, số lượng quan sát luôn nhỏ hơn đáng kể so với hai nguồn còn lại vì đơn vị đánh giá là nhóm học tập và GV chỉ thực hiện đánh giá đối với một số nhóm được lựa chọn sau mỗi bài học. Những khác biệt về đơn vị phân tích và quy mô quan sát này cần được lưu ý khi đọc và đối chiếu các kết quả trình bày trong các phần tiếp theo.

3.2. Xu hướng thay đổi của kết quả đánh giá năng lực giao tiếp và hợp tác trong môi trường số theo ba nguồn đánh giá

Bảng 2. Điểm trung bình của ba nguồn đánh giá theo bài học can thiệp (nhóm EXP)

Nguồn đánh giá	Bài học 1 M (SD)	Bài học 2 M (SD)	Bài học 3 M (SD)	Bài học 4 M (SD)
SA (Tự đánh giá)	2.08 (0.56)	2.09 (0.52)	2.27 (0.46)	2.41 (0.31)
PA (Đánh giá đồng đẳng)	2.15 (0.64)	2.08 (0.72)	2.44 (0.58)	2.46 (0.70)
TR (Đánh giá của GV)	1.88 (0.71)	2.03 (0.76)	2.38 (0.64)	2.37 (0.51)

Ghi chú. Thang đo: 0-3. Giá trị được trình bày dưới dạng M (SD). Cỡ mẫu (n) của từng nguồn đánh giá theo từng bài học được trình bày tại bảng 1.

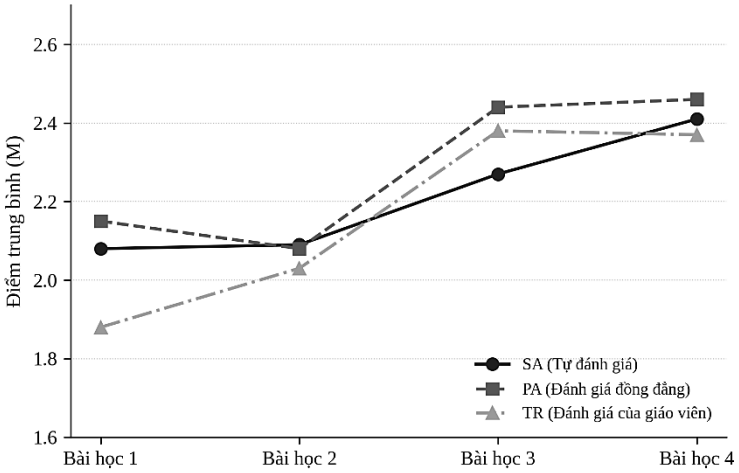
Kết quả tại bảng 2 và hình 1 cho thấy điểm trung bình của cả ba nguồn đánh giá đều có xu hướng tăng qua bốn bài học can thiệp. Trong hai bài học đầu tiên, mức thay đổi của từng nguồn chưa hoàn toàn đồng nhất: điểm tự đánh giá (SA) gần như ổn định, điểm đánh giá đồng đẳng (PA) giảm nhẹ, trong khi điểm đánh giá của GV (TR) tăng rõ hơn. Từ Bài học 2 sang Bài học 3, cả ba nguồn đều ghi nhận mức tăng rõ rệt, trước khi mức điểm trung bình thay đổi ít hơn ở Bài học 4; riêng đối với TR, xu hướng này cần được xem xét thận trọng do Bài học 4 chỉ còn 5 quan sát.

Xét theo từng nguồn đánh giá, điểm trung bình của SA tăng từ 2.08 ở Bài học 1 lên 2.41 ở Bài học 4. PA đạt mức 2.15 ở Bài học đầu tiên, giảm nhẹ ở Bài học 2 trước khi tăng lên 2.44 và 2.46 ở hai bài học cuối. Đối với TR, điểm trung bình tăng từ 1.88 lên 2.38 ở Bài học 3 và duy trì ở mức 2.37 tại Bài học 4.

So sánh giữa ba nguồn đánh giá cho thấy vẫn tồn tại sự khác biệt nhất định về mức điểm tuyệt đối trong suốt quá trình can thiệp. PA đạt điểm trung bình cao nhất trong số ba nguồn đánh giá ở ba trong bốn bài học (Bài học 1, 3 và 4), trong khi TR thấp hơn PA ở cả bốn bài học theo điểm trung bình quan sát được nhưng cao hơn SA tại Bài học 3 (TR = 2.38; SA = 2.27) - vì vậy, chênh lệch điểm giữa các nguồn không nên được hiểu là một nguồn đánh giá cao hơn, thấp hơn hoặc chính xác hơn nguồn khác về cùng một năng lực một cách hệ thống, mà phản ánh sự khác biệt về góc nhìn, phạm vi đo (PA chỉ đo ba trong sáu thành phần năng lực ở cấp cá nhân) và đơn vị phân tích (TR được ghi nhận ở cấp nhóm học tập). Xét cụ thể khoảng cách giữa PA và TR, mức chênh lệch giảm mạnh từ 0.27 điểm ở Bài học 1 xuống 0.05 điểm ở Bài học 2, sau đó dao động nhẹ ở mức 0.06 và 0.09 điểm tại Bài học 3 và Bài học 4, thay vì tiếp tục thu hẹp liên tục qua các bài học. Do TR ở Bài học 4 chỉ còn 5 quan sát, các so sánh liên quan đến giai đoạn cuối chỉ mang tính mô tả với độ ổn định hạn chế, chưa phải một xu hướng đã được xác lập chắc chắn.

3.3. Bàn luận

Sự nhất quán về xu hướng thay đổi của kết quả đánh giá năng lực GT&HTS được ghi nhận từ cả ba nguồn đánh giá cho thấy kết quả này phản ánh các xu hướng thay đổi nhìn chung tương đồng giữa các góc nhìn đánh giá khác nhau. Mặc dù ba nguồn sử dụng các chủ thể đánh giá và đơn vị phân tích khác nhau, cả tự đánh giá, đánh giá đồng đẳng và đánh giá của GV đều ghi nhận thời điểm chuyển biến rõ nhất từ Bài học 2 sang Bài học 3. Kết quả này trả lời trực tiếp câu hỏi nghiên cứu thứ nhất và cho thấy việc theo dõi đồng thời nhiều nguồn đánh giá giúp phản ánh xu hướng thay đổi của kết quả đánh giá năng lực từ nhiều góc nhìn khác nhau, thay vì chỉ dựa vào một nguồn đánh giá đơn lẻ.



Hình 1. Điểm trung bình của ba nguồn đánh giá theo bài học can thiệp (nhóm EXP)

Ghi chú: Cỡ mẫu (n) theo từng bài học và nguồn đánh giá được trình bày tại bảng 1; do cỡ mẫu thay đổi qua các thời điểm, xu hướng trong hình phản ánh dữ liệu quan sát được tại mỗi thời điểm, không phải cùng một nhóm HS cố định.

Bên cạnh sự thống nhất về xu hướng, nghiên cứu cũng ghi nhận sự khác biệt nhất định về mức điểm tuyệt đối giữa ba nguồn đánh giá. Đánh giá của GV thấp hơn tự đánh giá và đánh giá đồng đẳng ở phần lớn các bài học, ngoại trừ Bài học 3, khi điểm đánh giá của GV (2.38) cao hơn điểm tự đánh giá (2.27); khác biệt này rõ nhất trong giai đoạn đầu của quá trình can thiệp. Kết quả này có xu hướng phù hợp với phát hiện của nghiên cứu trước của nhóm tác giả (Cao Hồng Huệ và cộng sự, 2026), dù không hoàn toàn nhất quán ở mọi bài học. Đáng chú ý, mặc dù mức độ khác biệt giữa các nguồn chưa giảm một cách đồng đều qua từng bài học, khoảng cách giữa các nguồn ở giai đoạn cuối nhìn chung nhỏ hơn so với giai đoạn đầu của quá trình can thiệp. Nhìn chung, ba nguồn đánh giá cho thấy điểm giống nhau về xu hướng thay đổi theo thời gian nhưng khác nhau về mức điểm tuyệt đối, đặc biệt ở giai đoạn đầu can thiệp; sự tương đồng về hình dạng đường biểu diễn giữa ba nguồn phản ánh đặc điểm mô tả của dữ liệu, không hàm ý các nguồn đạt được sự đồng thuận hoặc giá trị hội tụ về mặt đo lường. Kết quả này trả lời trực tiếp câu hỏi nghiên cứu thứ hai và cho thấy việc phối hợp đánh giá đa nguồn trong dạy học Tin học ở tiểu học cần tính đến cả điểm giống và khác giữa các nguồn theo từng giai đoạn của quá trình can thiệp, thay vì giả định các nguồn phản ánh năng lực ở cùng một mức độ tại mọi thời điểm.

Những phát hiện này bổ sung thêm bằng chứng thực nghiệm cho nghiên cứu về đánh giá năng lực GT&HTS ở HS tiểu học. Kết quả cho thấy xu hướng thay đổi của kết quả đánh giá năng lực không chỉ được ghi nhận nhất quán giữa các nguồn đánh giá mà còn phản ánh sự khác biệt về mức điểm giữa các nguồn theo từng giai đoạn của quá trình can thiệp. Đây là khía cạnh mà các nghiên cứu trước đây, vốn chủ yếu dựa trên một nguồn đánh giá hoặc chỉ khảo sát tại một thời điểm, chưa có điều kiện xem xét. Theo đó, kết quả của nghiên cứu này bổ sung một chiều bằng chứng thực nghiệm mà tài liệu hiện có về đánh giá năng lực GT&HTS ở HS tiểu học còn thiếu, đặc biệt là bằng chứng về cách các nguồn đánh giá vận động cùng nhau theo thời gian trong bối cảnh dạy học môn Tin học.

Bên cạnh cách diễn giải xu hướng tăng điểm như một dấu hiệu phát triển năng lực, cũng cần cân nhắc một số khả năng giải thích thay thế mà nghiên cứu này chưa đo lường trực tiếp: (1) Xu hướng tăng điểm quan sát được có thể phản ánh một phần hiệu ứng làm quen (familiarity effect), khi HS và GV trở nên quen thuộc hơn với công cụ và quy trình đánh giá qua các bài học liên tiếp; (2) Do nội dung chuyên môn của các bài học khác nhau giữa các khối lớp, sự khác biệt về độ khó hoặc đặc điểm nhiệm vụ giữa các bài học cũng có thể góp phần vào sự thay đổi điểm số quan sát được, độc lập với sự phát triển năng lực thực chất; (3) Như đã nêu ở phần Phương pháp, thành phần mẫu tham gia không hoàn toàn giống nhau giữa các bài học, nên một phần biến động điểm số theo thời gian có thể phản ánh sự thay đổi thành phần mẫu hơn là một xu hướng phát triển thống nhất. Mặc dù không thể tách bạch định lượng đóng góp của từng khả năng này với dữ liệu hiện có, các khả năng giải thích thay thế nêu trên cần được xem là giới hạn trong việc diễn giải xu hướng tăng điểm quan sát được.

4. Kết luận và bình luận

Nghiên cứu đã mô tả và đối chiếu xu hướng thay đổi của kết quả đánh giá năng lực GT&HTS của HS tiểu học thông qua ba nguồn đánh giá độc lập gồm tự đánh giá, đánh giá đồng đẳng và đánh giá của GV trong bối cảnh thực nghiệm sư phạm. Kết quả cho thấy xu hướng này được ghi nhận nhất quán giữa ba nguồn đánh giá, đồng thời phản ánh những khác biệt nhất định về mức điểm tuyệt đối giữa các nguồn ở từng thời điểm của quá trình can thiệp. Những phát hiện này bổ sung bằng chứng thực nghiệm về tính khả thi của việc sử dụng đánh giá đa nguồn để theo dõi xu hướng thay đổi của kết quả đánh giá năng lực GT&HTS của HS tiểu học trong dạy học Tin học.

Về mặt thực tiễn, mức độ khác biệt giữa các nguồn đánh giá lớn hơn ở giai đoạn đầu của quá trình can thiệp cho thấy GV cần thận trọng khi diễn giải kết quả đánh giá đa nguồn trong những bài học đầu tiên. Khác biệt về mức điểm giữa các nguồn giảm mạnh từ Bài học 1 sang Bài học 2, sau đó dao động nhẹ ở hai bài học cuối thay vì tiếp tục thu hẹp liên tục qua các giai đoạn sau; nhận định này cũng cần được xem xét cùng với sự suy giảm cỡ mẫu ở giai đoạn cuối, đặc biệt đối với đánh giá của GV, qua đó cung cấp thêm cơ sở tham khảo cho việc vận dụng phối hợp nhiều nguồn đánh giá trong quá trình dạy học môn Tin học ở tiểu học. Trong thực tiễn dạy học, kết quả của từng nguồn đánh giá nên được xem là thông tin bổ sung cho nhau hơn là thay thế cho nhau. Tự đánh giá có thể giúp HS tự nhìn nhận quá trình học tập của bản thân, đánh giá đồng đẳng cung cấp thêm góc nhìn từ bạn học, trong khi đánh giá của GV đóng vai trò như một nguồn tham chiếu chuyên môn. Việc phối hợp các nguồn đánh giá này có thể hỗ trợ GV hình thành cái nhìn toàn diện hơn về kết quả đánh giá của HS trong quá trình dạy học môn Tin học ở tiểu học.

Nghiên cứu cũng có một số hạn chế cần được cân nhắc khi diễn giải kết quả: (1) Do mục tiêu tập trung vào việc mô tả xu hướng thay đổi của kết quả đánh giá năng lực theo thời gian, nghiên cứu chủ yếu sử dụng thống kê mô tả mà chưa thực hiện các kiểm định suy luận thống kê hoặc tính toán độ lớn hiệu ứng; vì vậy, các kết quả phản ánh đặc điểm của dữ liệu quan sát được hơn là bằng chứng suy luận về sự khác biệt giữa các thời điểm; (2) Ba nguồn đánh giá sử dụng hai đơn vị phân tích khác nhau, với tự đánh giá và đánh giá đồng đẳng được ghi nhận ở cấp độ cá nhân HS, trong khi đánh giá của GV được thực hiện ở cấp độ nhóm học tập, điều này cần được lưu ý khi so sánh trực tiếp giữa các nguồn; (3) Do số lượng trường và lớp tham gia đầy đủ cả bốn bài học không hoàn toàn giống nhau, thành phần mẫu ở các thời điểm sau có thể khác với thời điểm đầu, do đó, xu hướng thay đổi của kết quả đánh giá quan sát được nên được hiểu là đặc điểm của dữ liệu thu được qua từng thời điểm hơn là quỹ đạo phát triển của cùng một nhóm HS cố định; cỡ mẫu của đánh giá GV giảm đặc biệt mạnh ở bài học cuối (còn 5 quan sát), làm giảm độ ổn định của các ước lượng tại thời điểm này; (4) Phiếu đánh giá đồng đẳng chỉ bao phủ ba trong sáu thành phần của khung năng lực, nên phạm vi đo lường chưa hoàn toàn tương đương với hai công cụ còn lại. Đồng thời, phiếu này chưa được thẩm định bởi chuyên gia độc lập hoặc thử nghiệm pilot trước khi áp dụng; vì vậy, kết quả thu được từ nguồn đánh giá đồng đẳng cần được diễn giải thận trọng và cần tiếp tục được kiểm chứng trong các nghiên cứu tiếp theo; (5) Tại một số lớp, dữ liệu đánh giá đồng đẳng chưa được thu thập đầy đủ do điều kiện triển khai thực tế tại các trường tham gia, dẫn đến số lượng quan sát của nguồn đánh giá này thấp hơn so với tự đánh giá ở một số bài học. Đây là dữ liệu thiếu phát sinh trong quá trình thu thập dữ liệu tại các trường tham gia nghiên cứu, không phải do sai sót trong quá trình xử lí dữ liệu.

Từ những hạn chế trên, nghiên cứu tiếp theo có thể áp dụng các phương pháp phân tích suy luận phù hợp với dữ liệu đo lặp nhằm kiểm chứng thống kê đối với các xu hướng đã được mô tả trong nghiên cứu này. Đồng thời, việc phát triển các công cụ đánh giá có phạm vi đo lường tương đương giữa các nguồn, đồng nhất đơn vị phân tích và duy trì cỡ mẫu ổn định hơn qua các thời điểm sẽ góp phần nâng cao khả năng so sánh cũng như độ tin cậy của đánh giá đa nguồn. Bên cạnh đó, việc mở rộng phiếu đánh giá đồng đẳng để bao phủ đầy đủ sáu thành phần của khung năng lực cũng là một hướng nghiên cứu cần được xem xét nhằm hoàn thiện hệ thống đánh giá năng lực GT&HTS của HS tiểu học trong dạy học Tin học.

Tuyên bố về vai trò của các tác giả: *Cao Hồng Huệ: Xây dựng ý tưởng nghiên cứu, Thiết kế phương pháp nghiên cứu, Thực hiện nghiên cứu và thu thập dữ liệu, Quản lí và xử lí dữ liệu, Phân tích dữ liệu, Trực quan hóa dữ liệu, Viết bản thảo ban đầu. Đào Thái Lai, Nguyễn Thị Thanh Huyền: Giám sát và hướng dẫn, Thẩm định và kiểm chứng kết quả. Cao Hồng Huệ, Đào Thái Lai, Nguyễn Thị Thanh Huyền: Chỉnh sửa và phản biện bản thảo.*
Tuyên bố về GenAI và Quyền tác giả: *Trong quá trình chuẩn bị bản thảo này, các tác giả đã sử dụng ChatGPT (OpenAI) và Claude (Anthropic) nhằm hỗ trợ biên tập ngôn ngữ, rà soát cấu trúc học thuật, cải thiện diễn đạt và hỗ trợ tổng hợp tài liệu tham khảo. Toàn bộ nội dung do các công cụ AI hỗ trợ đã được các tác giả kiểm tra, chỉnh sửa và xác minh.*
Tuyên bố về xung đột lợi ích: *Các tác giả tuyên bố không có xung đột lợi ích.*
Thông tin tài trợ: *Nghiên cứu này được tài trợ từ nguồn kinh phí Khoa học Công nghệ của Trường Đại học Sư phạm Hà Nội 2 cho Đề tài Khoa học, mã số: HPU2.2024-UT-04.*
Lời cảm ơn: *Nhóm tác giả trân trọng cảm ơn Trường Đại học Sư phạm Hà Nội 2 đã tài trợ kinh phí thực hiện nghiên cứu thông qua Đề tài Khoa học, mã số HPU2.2024-UT-04.*

Tài liệu tham khảo

Cao, H. H., Dao, L. T., Tran, T., & Nguyen, H. T. T. (2025). Communication and collaboration competence within the digital competence framework: A bibliometric analysis. *International Journal of Evaluation and Research in Education*, *14*(4), 2652-2665. https://doi.org/10.11591/ijere.v14i4.32183

Cao Hồng Huệ, Đào Thái Lai, Nguyễn Thị Thanh Huyền, Trần Thị Thanh (2025). Xây dựng khung năng lực giao tiếp và hợp tác thuộc khung năng lực số của học sinh tiểu học trong dạy học Tin học. *Tạp chí Giáo dục*, *25*(19), 41-46. https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/4160

Cao Hồng Huệ, Đào Thái Lai, Nguyễn Thị Thanh Huyền (2026). Năng lực giao tiếp và hợp tác trong môi trường số của học sinh tiểu học: Kiểm định thang đo và đối chiếu giữa tự đánh giá của học sinh và đánh giá của giáo viên. *Tạp chí Khoa học Giáo dục Việt Nam*, *22*(06), 62-71.

Černochová, M., Voňková, H., Štípek, J., & Černá, P. (2020). How do learners perceive and evaluate their digital skills? In I. Management Association (Ed.), *Learning and Performance Assessment: Concepts, Methodologies, Tools, and Applications* (pp. 1453-1464). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-7998-0420-8.ch067

Claes, E., Golsteyn, B. H. H., de Leeuw, S., & de Wolf, I. (2026). Primary school skill development: From first to sixth grade. *Journal of Human Capital*, *20*(2), 261-286. https://doi.org/10.1086/740863

Devaki, V. (2024). Exploring the impact of innovative assessment methods on learning outcomes: A review of project-based assessments, portfolios, peer assessment, and self-assessment. In G. Zeybek (Ed.), *Policy development, curriculum design, and administration of language education* (pp. 343-374). IGI Global. https://doi.org/10.4018/979-8-3693-3645-8.ch015

Godaert, E., Aesaert, K., Voogt, J., & van Braak, J. (2022). Assessment of students' digital competences in primary school: A systematic review. *Education and Information Technologies*, *27*(7), 9953-10011. https://doi.org/10.1007/s10639-022-11020-9

León-Sandobal, S., Rivas-Soriano, V., & Gonzales-Macavilca, M. (2025). Digital generation? Digital competencies in primary school students in Lima-Norte (Peru). *Lecture Notes in Networks and Systems*, *1348*, 376-387. https://doi.org/10.1007/978-3-031-87701-8_27

Nguyen, T. Q., Anh Ngoc, P. T., Phuong, H. A., Duy, D. P. T., Hiep, P. C., McClelland, R., & Noroozi, O. (2024). Digital competence of Vietnamese citizens: An application of DigComp framework and the role of individual factors. *Education and Information Technologies*, *29*(15), 19267-19298. https://doi.org/10.1007/s10639-024-12585-3

Porat, E., Blau, I., & Barak, A. (2018). Measuring digital literacies: Junior high-school students' perceived competencies versus actual performance. *Computers & Education*, *126*, 23-36. https://doi.org/10.1016/j.compedu.2018.06.030

Widodo, G. S., Ekohariadi, Anistyasari, Y., Sabtiawan, W. B., Rijanto, T., Zainuddin, A., Kristiandri, D., Hidayah, L., & Pratama, S. A. (2025). Collaborative assessment in education: A systematic review of practices and impact in schools (2014-2024). *Multidisciplinary Reviews*, *9*(5), Article 2026241. https://doi.org/10.31893/multirev.2026241