

SUY LUẬN THỐNG KÊ CỦA SINH VIÊN THỂ HIỆN QUA HOẠT ĐỘNG KHẢO SÁT THỐNG KÊ

Nguyễn Thị Tân An¹, Huỳnh Trí Bình²

Tóm tắt: Trong nghiên cứu này, chúng tôi tìm hiểu suy luận thống kê của sinh viên sư phạm Toán thể hiện thông qua một khảo sát thống kê, từ đó đưa ra những đề xuất đối việc dạy học Thống kê giúp nâng cao suy luận thống kê mà sinh viên sư phạm ngành Toán cần được rèn luyện trong quá trình học tập ở trường đại học. Nghiên cứu sử dụng phương pháp thực nghiệm để quan sát và phân tích suy luận thống kê của sinh viên. Kết quả cho thấy mặc dù hai phần ba số nhóm tham gia thực nghiệm thuộc mức độ 4/5 của thang đo SOLO nhưng quá trình phân tích định tính bộc lộ những khiếm khuyết trong suy luận thống kê của sinh viên cũng như chương trình giảng dạy Thống kê ở bậc Phổ thông hiện tại. Nghiên cứu cũng chứng tỏ, việc sử dụng quy trình khảo sát thống kê tạo điều kiện cho sinh viên thể hiện suy luận thống kê và đem lại hứng thú học tập cho sinh viên, giúp các em hiểu rõ hơn ý nghĩa của Thống kê trong thực tế, góp phần phát triển khả năng vận dụng kiến thức Thống kê vào thực tiễn.

Từ khóa: Quy trình khảo sát thống kê, suy luận thống kê, thống kê.

1. ĐẶT VẤN ĐỀ

Trong xã hội hiện nay, các tuyên bố và báo cáo nghiên cứu xuất hiện thường xuyên trên các phương tiện truyền thông, vì vậy hiểu biết thống kê là rất cần thiết để đưa ra những quyết định đúng đắn thay vì dựa vào cảm xúc và niềm tin [1]. Do tầm quan trọng của thống kê trong cuộc sống hằng ngày và nơi làm việc nên hiểu biết, suy luận, và tư duy thống kê đang trở thành những năng lực quan trọng cần phát triển cho học sinh (HS) ở trường phổ thông tại nhiều quốc gia trên thế giới [2]. Garfield và Ben-Zvi [3] cho rằng, mặc dù việc

¹ Trường Đại học Sư phạm Huế, Email: nguyenthitanan@dhsphue.edu.vn.

² Trường Đại học Quy Nhơn, Email: huynhtribinh@qnu.edu.vn.

cải cách trong học tập và giảng dạy Thống kê được đề cập rộng rãi nhưng dạy học Thống kê vẫn được coi là thách thức bởi vì giáo dục (GD) Thống kê trong trường học thường tập trung vào việc phát triển kỹ năng tính toán, điều đó là không đủ để HS suy luận hoặc tư duy thống kê [4].

Hiện tại, trong chương trình Thống kê ở trường phổ thông (lớp 7 và 10), HS thường được cung cấp dữ liệu sẵn để tính toán hoặc chứng minh một kết quả được xác định trước. HS hầu như không có trải nghiệm về suy luận thống kê cũng như ra quyết định cần thiết để lập kế hoạch thu thập dữ liệu, ghi chép dữ liệu và phân tích dữ liệu để đưa ra những kết luận thống kê phù hợp. Tuy nhiên, trong Chương trình GD phổ thông (GDPT) môn Toán 2018, Thống kê đã được chú trọng và trở thành một trong ba mạch kiến thức chính, xuyên suốt từ lớp 2 đến lớp 12, nhằm hình thành sự hiểu biết về vai trò của Thống kê cho HS cũng như giúp các em biết áp dụng tư duy thống kê để phân tích dữ liệu, giải quyết một số vấn đề thực tiễn đơn giản gắn với một số yếu tố Thống kê và xác suất, góp phần tăng cường tính ứng dụng và giá trị thiết thực của GD Toán học. Ngoài ra, các bước của quy trình khảo sát thống kê cũng được lồng ghép trong chương trình. Vì vậy, việc đào tạo giáo viên (GV) ở các trường sư phạm cũng cần có những thay đổi để phù hợp với chương trình GDPT mới, cũng như đáp ứng được yêu cầu dạy học trong giai đoạn mới. Trong nghiên cứu này, chúng tôi muốn tìm hiểu suy luận thống kê của sinh viên (SV) Sư phạm Toán thể hiện như thế nào thông qua một khảo sát thống kê đơn giản, từ đó có những đề xuất sư phạm liên quan đến việc phát triển suy luận thống kê, tư duy thống kê của GV Toán tương lai. *Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số 503.01-2020.313.*

2. NỘI DUNG NGHIÊN CỨU

2.1. Sự khác biệt giữa Thống kê và Toán học

Ở trường phổ thông, Thống kê được đặt trong chương trình giảng dạy của môn Toán, tuy nhiên vẫn có sự khác biệt giữa Toán học và Thống kê (Moore và Cobb, 2000). Trong Thống kê, chúng ta sử dụng các công cụ Toán học để giải quyết vấn đề, nhưng lại dựa vào dữ liệu và bối cảnh để suy luận.

Nếu như Toán học cố gắng loại bỏ bối cảnh để có thể hướng tới sự trừu tượng hóa, thì đối với Thống kê bối cảnh cung cấp ý nghĩa cho dữ liệu, chi phối cách thu thập dữ liệu, phân tích và giải thích kết quả [5]. Chúng ta không thể hiểu được vấn đề Thống kê nếu không biết bối cảnh xung quanh dữ liệu. Bên cạnh đó, Toán học có thể được nghiên cứu và thực hiện mà không liên quan đến dữ liệu, nhưng trong Thống kê, dữ liệu rất quan trọng và chiến lược thu thập, phân tích dữ liệu sẽ ảnh hưởng đến kết luận được rút ra [6]. Tính không xác định hoặc “lộn xộn” của dữ liệu trong Thống kê khác với bản chất hữu hạn, chính xác đặc trưng cho Toán học.

Một vấn đề khác biệt nữa giữa Toán học và Thống kê là phép đo. Trong Toán học, để đo mỗi đối tượng sẽ có phép đo, công thức đo, đơn vị đo cụ thể. Tuy nhiên, đối với Thống kê, có thể có nhiều phép đo khác nhau để đo cùng một đối tượng, và để rút ra kết luận có giá trị từ dữ liệu thì điều này phụ thuộc rất nhiều vào việc lựa chọn phép đo. Ví dụ, để đo lường trí thông minh hoặc nhịp sống của một thành phố thì không có một phương pháp cụ thể nào, các nhà nghiên cứu và nhà Thống kê học phải quyết định cách để đo lường tốt nhất những gì đang nghiên cứu và thường thực hiện theo những cách khác nhau [6].

Ngoài ra, tính biến thiên và không chắc chắn của các kết luận thống kê là một điểm khác biệt lớn giữa Thống kê và Toán học. Có thể có nhiều kết luận khác nhau nhưng hợp lý khi phân tích cùng một dữ liệu, điều này phần lớn là do việc giải thích bối cảnh, phương pháp thu thập và phân tích dữ liệu, cũng như bản chất biến thiên của các vấn đề thống kê. Ví dụ: “Các GV trong trường của bạn bao nhiêu tuổi?” là một câu hỏi thống kê. Để trả lời câu hỏi này, chúng ta cần quyết định lấy dữ liệu từ đâu, phép đo gì, chọn công cụ nào để tóm tắt số liệu thống kê thích hợp và vẽ loại biểu đồ nào để trả lời câu hỏi. Ngược lại, nếu cho một tập dữ liệu về tuổi của GV và yêu cầu HS tìm giá trị trung bình của tập dữ liệu thì đó không phải là một câu hỏi thống kê, vì câu trả lời là một con số duy nhất được tìm thấy bằng một thuật toán.

Chính vì những khác biệt trên mà Thống kê đòi hỏi những cách suy luận khác hẳn với suy luận Toán học [5]. Suy luận thống kê phụ thuộc vào dữ liệu và bối cảnh (Chance, 2002), và suy luận Toán học chỉ được sử dụng hỗ trợ giải quyết vấn đề thống kê sau khi đã xác định được câu hỏi điều tra, lựa chọn thiết kế phù hợp để thu thập dữ liệu và mô hình thích hợp để phân tích dữ liệu.

2.2. Suy luận thống kê

Suy luận thống kê có thể được hiểu là cách sử dụng các kiến thức thống kê để suy luận và hiểu thông tin thống kê [7]. Theo Garfield và Ben-Zvi [8], suy luận thống kê là khả năng:

- Giải thích dựa trên dữ liệu, biểu diễn dữ liệu, và tóm tắt thống kê dữ liệu.
- Kết nối các khái niệm, các ý tưởng thống kê để đưa ra kết luận có ý nghĩa thống kê.
- Hiểu và giải thích các quy trình thống kê, lí giải các kết quả thống kê.

DeIMas [9] cho rằng suy luận thống kê được thể hiện khi một người có thể giải thích tại sao hoặc làm thế nào để đưa ra kết quả, tại sao kết luận là hợp lí, tại sao lại thích hợp để lựa chọn một mô hình hoặc biểu diễn thống kê trong một bối cảnh cụ thể.

Cơ sở của suy luận thống kê là hiểu các khái niệm quan trọng như phân phối, tâm, tính biến thiên, độ biến động, sự không chắc chắn, sự ngẫu nhiên, và mẫu...

Các loại suy luận thống kê sau đây được chú trọng để phát triển cho HS khi học về thống kê trong nhà trường [7]:

- Suy luận liên quan đến đặc tính của mẫu số liệu: Nhận ra một mẫu số liệu là định lượng hay định tính, rời rạc hay liên tục; Biết được loại bảng, biểu đồ, hoặc phương pháp thống kê nào phù hợp tương ứng với mỗi loại mẫu số liệu.

- Suy luận liên quan đến trình bày mẫu số liệu: Biết cách biểu diễn một mẫu số liệu bằng biểu đồ phù hợp; Biết đọc và giải thích các loại biểu đồ.

- Suy luận liên quan đến các phép đo thống kê: Biết cách tính các tham số định tâm (số trung bình, trung vị, mốt), các tham số đo độ phân tán (biên độ, độ lệch trung bình, phương sai, độ lệch chuẩn) của mẫu số liệu; Biết cách sử dụng các số đặc trưng nào là phù hợp cho từng trường hợp cụ thể.

- Suy luận liên quan đến xác suất: Hiểu và sử dụng những ý tưởng về ngẫu nhiên, cơ hội, khả năng để đánh giá về những biến cố không chắc chắn; Biết sử dụng phương pháp thích hợp để xác định khả năng xảy ra của các biến cố khác nhau.

- Suy luận về mẫu: Biết trong những tình huống cụ thể nào thì việc điều tra toàn bộ là không khả thi mà phải điều tra mẫu; Biết cách chọn mẫu từ toàn bộ tập hợp đơn vị điều tra như thế nào là hợp lí (mẫu được chọn phải ngẫu nhiên, đủ lớn và đại diện cho toàn bộ tập hợp đơn vị điều tra); Biết rút ra các kết luận cho tổng thể từ việc phân tích các thông tin thu được trên mẫu đó.

- Suy luận về mối quan hệ giữa hai biến số: Biết đánh giá và giải thích mối quan hệ giữa hai biến số; Biết phân tích và giải thích bằng hai chiều hay biểu đồ phân tán khi xem xét mối quan hệ giữa hai biến số.

Dạy học để phát triển suy luận thống kê

Theo lí thuyết kiến tạo, mọi người học không phải bằng cách tiếp nhận kiến thức mà bằng cách xây dựng kiến thức mới dựa trên kiến thức và niềm tin hiện có, kiến thức cũ không mất đi mà nó được tích hợp và điều chỉnh để phù hợp với kiến thức mới [10]. Vì vậy, hoạt động giảng dạy tốt sẽ bao gồm việc thiết kế môi trường học tập để kích thích HS xây dựng kiến thức, tạo cơ hội để các em phản ánh, cũng như thảo luận và thực hành, từ đó phát triển việc hiểu sâu sắc về kiến thức được học, có thể áp dụng kiến thức đó ở các lớp học tiếp theo hoặc trong thế giới thực. Dựa trên nền tảng của lí thuyết kiến tạo, Cobb và McClain [11] đã đề xuất 6 nguyên tắc dạy học giúp phát triển suy luận thống kê sau đây:

- Tập trung vào việc phát triển các ý tưởng thống kê hơn là trình bày công thức và quá trình.

- Sử dụng dữ liệu thực và có động lực để thúc đẩy HS đưa ra, kiểm nghiệm các phỏng đoán.

- Sử dụng các hoạt động học tập để hỗ trợ sự phát triển suy luận của HS.

- Tích hợp việc sử dụng các công cụ công nghệ phù hợp cho phép SV kiểm tra các phỏng đoán, khám phá và phân tích dữ liệu cũng như phát triển suy luận thống kê.

- Khuyến khích các lập luận thống kê và trao đổi tập trung vào các ý tưởng thống kê quan trọng.

- Sử dụng đánh giá để tìm hiểu những gì HS biết và để theo dõi sự phát triển của việc học Thống kê, cũng như để đánh giá kế hoạch và tiến trình giảng dạy.

2.3. Quy trình khảo sát thống kê

Để phát triển suy luận thống kê của HS, DelMas [9] cho rằng, các em nên được tham gia vào quá trình thu thập và khám phá dữ liệu, thảo luận về cách dữ liệu được tạo ra, lựa chọn các tóm tắt thống kê thích hợp và rút ra kết luận dựa trên bằng chứng. Wild và Pfannkuch [12] đã mô tả quá trình mà các nhà thống kê sử dụng trong thực hành khảo sát dựa trên dữ liệu, gồm 5 bước từ việc xây dựng vấn đề đến kết luận. Quá trình này đã tạo cơ sở cho nhiều nghiên cứu trong GD Toán học bằng cách tập trung vào bản chất và sự phát triển của hiểu biết, suy luận và tư duy thống kê [4].

Bước 1 (Vấn đề): Bước đặt vấn đề là rất cần thiết trong quy trình khảo sát thống kê vì vấn đề khảo sát đóng vai trò kết nối và tạo động lực thúc đẩy khảo sát. Các vấn đề đặt ra để tiến hành khảo sát thống kê cần thú vị, thách thức và phù hợp với đối tượng người học. Hơn nữa, vấn đề cần có bản chất thống kê, nghĩa là để trả lời được câu hỏi đặt ra người học cần thu thập và giải thích dữ liệu. Ngoài ra, dữ liệu nên cung cấp đủ độ phức tạp để tạo ra các kết quả thú vị và GV nên sử dụng các câu hỏi cho phép người học tự định hướng quá trình giải quyết, điều này sẽ dẫn dắt các cuộc khảo sát liên quan đến các lĩnh vực thống kê khác nhau.

Bước 2 (Lập kế hoạch): Tư duy thống kê ở giai đoạn này liên quan đến việc xem xét chủ đề và bối cảnh, sử dụng biến nào, dữ liệu nào cần lấy và cách lấy dữ liệu đại diện. Các câu hỏi có thể được đặt ra để hỗ trợ việc lập kế hoạch, bất kể dữ liệu sẽ được thu thập, lựa chọn hay cung cấp, ví dụ như: Chúng ta muốn tìm hiểu về điều gì? Chúng ta có thể tìm được gì? Chúng ta có thể đo lường hoặc quan sát những gì? Làm thế nào chúng ta có thể thu thập dữ liệu?... Như vậy, giai đoạn lập kế hoạch của quy trình khảo sát thống kê bao gồm xác định các biến số, đối tượng nghiên cứu và xem xét tính thực tiễn liên quan đến vấn đề đặt ra.

Bước 3 (Dữ liệu): Dữ liệu có thể được tạo ra thông qua các cuộc khảo sát, thí nghiệm, quan sát hoặc từ các bộ dữ liệu đã có sẵn trên Internet hoặc

các nguồn khác. Giai đoạn này bao gồm việc thu thập, xử lý và làm sạch dữ liệu. Khi thu thập và chuẩn bị dữ liệu để phân tích, cần quan tâm đến việc biểu diễn vấn đề và các biến, cùng với các khó khăn có thể gặp khi đo lường. Trong trường hợp dữ liệu được cung cấp sẵn, quá trình thu thập dữ liệu vẫn có thể được thảo luận để người học hiểu được sự cần thiết của dữ liệu đại diện và cách thức thu thập dữ liệu.

Bước 4 (Phân tích/xử lý dữ liệu): Giai đoạn này chủ yếu đề cập đến việc lựa chọn và sử dụng các biểu diễn, tóm tắt dữ liệu để khám phá, nhận xét về đặc điểm của dữ liệu. Ở mức độ cao hơn, liên quan đến việc khảo sát sự biến thiên (trong tập dữ liệu, giữa các nhóm dữ liệu, giữa các tập dữ liệu có bối cảnh giống nhau hoặc tương tự), suy luận với các mô hình thống kê và kết hợp số liệu thống kê với bối cảnh.

Bước 5 (Kết luận/thảo luận): Điều quan trọng trong tư duy thống kê là nhấn mạnh vào việc bình luận, giải thích và thảo luận, chứ không chỉ giới hạn trong việc “trả lời câu hỏi”. Có thể có nhiều câu trả lời được đưa ra, nhưng nên tập trung vào những nhận xét và diễn giải phù hợp. Nên phân biệt giữa những gì dữ liệu cho chúng ta biết và những gì có thể là suy đoán, đưa ra kết luận dựa trên trực giác về ngữ cảnh.

2.4. Phương pháp nghiên cứu

Nghiên cứu được thực hiện với 57 SV Sư phạm ngành Toán năm 3. Tại thời điểm này, SV chưa tham gia học phần nào về Thống kê ở bậc ĐH, kiến thức thống kê hiện có của các em là những gì được học ở bậc Phổ thông. Chúng tôi chia lớp thành 9 nhóm, mỗi nhóm từ 6 đến 7 SV để trao đổi, thảo luận về các câu hỏi đặt ra.

Phương pháp nghiên cứu chúng tôi thực hiện ở nghiên cứu này là phương pháp thực nghiệm để có thể quan sát và phân tích suy luận thống kê của người học khi thực hiện quy trình khảo sát thống kê. Trên cơ sở các nguyên tắc dạy học của Cobb và McClain [11] đã giới thiệu ở mục 2.3, chúng tôi thiết kế tình huống thực nghiệm như sau:

- Lớp học bắt đầu với việc ôn lại những kiến thức thống kê mà SV đã được học ở bậc Phổ thông: Các bảng phân bố tần số, tần suất, các loại biểu

đồ, tham số định tâm (trung bình, trung vị, môđ), phương sai và độ lệch chuẩn. Vấn đề được đưa ra cho các SV (xem Hình 1):

Ngày 26 tháng 12 năm 1961, Thủ tướng Phạm Văn Đồng kí ban hành quyết định về việc “sinh đẻ có hướng dẫn” và Việt Nam trở thành một trong những quốc gia đầu tiên ở châu Á thực hiện chương trình dân số kế hoạch hóa gia đình. Giai đoạn năm 1991 - 2000, Hội nghị Ban chấp hành Trung ương khóa 5 đề ra mục tiêu “thực hiện gia đình ít con, khỏe mạnh, tạo điều kiện để có cuộc sống ấm no, hạnh phúc”, “mỗi gia đình chỉ có 1 hoặc 2 con”. Nhờ thực hiện tốt kế hoạch hóa gia đình nên vào năm 2007, Việt Nam chính thức bước vào thời kì dân số vàng, đây là cơ hội lớn để phát triển đất nước. Trong nhiều năm qua, nước ta đã duy trì được mức sinh cân bằng, trung bình mỗi phụ nữ trong độ tuổi sinh đẻ sinh 2,1 con (tỉ lệ sinh là 2,1). Tuy nhiên có sự chênh lệch giữa tỉ lệ sinh ở thành thị và nông thôn. Ví dụ như ở Thành phố Hồ Chí Minh, tỉ lệ sinh đầu năm 2021 chỉ là 1,53. Các câu hỏi được đặt ra:

Câu hỏi 1. Theo em, tỉ lệ sinh của gia đình các thành viên trong lớp so với tỉ lệ sinh của Việt Nam như thế nào? Giải thích.

Câu hỏi 2. Có mối quan hệ nào giữa số con trong gia đình của các thành viên trong lớp và độ tuổi của mẹ không? Giải thích.

Câu hỏi 3. Có mối quan hệ nào giữa số con trong gia đình của các thành viên trong lớp và nơi sinh sống của gia đình không? Giải thích.

Hình 1: Vấn đề khảo sát thống kê

- Ở giai đoạn lập kế hoạch và thu thập dữ liệu, giảng viên đề nghị các nhóm thảo luận “Em muốn tìm hiểu về điều gì? Những dữ liệu nào em cần thu thập? Làm thế nào để thu thập?”. Sau khi thảo luận, các nhóm thống nhất sử dụng phiếu khảo sát trực tuyến (google sheet) để thu thập dữ liệu của các thành viên trong lớp, bao gồm: Số con trong mỗi gia đình, tuổi của mẹ, nơi sinh sống của gia đình (thành thị / không thành thị). Sau khi các thành viên trong lớp trả lời phiếu khảo sát, giảng viên xuất dữ liệu dưới dạng file excel và gửi cho các nhóm.

- Giảng viên khuyến khích SV đưa ra các phỏng đoán cho các câu hỏi trên. Sau đó, yêu cầu SV sử dụng công cụ thống kê để kiểm tra các phỏng đoán, giải thích lập luận và chứng minh câu trả lời của mình. Các nhóm có thể sử dụng internet, các phần mềm để hỗ trợ cho quá trình trả lời các câu hỏi, bài làm được đánh máy và nộp lại cho giảng viên.

Trong nghiên cứu này, chúng tôi sử dụng phương pháp kết hợp giữa nghiên cứu định tính và nghiên cứu định lượng để phân tích dữ liệu. Dữ liệu

định lượng gồm kết quả đánh giá câu trả lời của SV theo phân loại SOLO của Biggs và Collis [13], gồm 5 mức độ xếp từ thấp đến cao như sau:

- *Mức độ 1* (Tiền cấu trúc): Người học bị xao nhãng bởi những yếu tố không liên quan, chưa nắm được ý tưởng. Kiến thức còn rời rạc, nông cạn.

- *Mức độ 2* (Đơn cấu trúc): Người học đã tập trung vào vấn đề và chọn một yếu tố nào đó của nhiệm vụ để giải quyết, chưa có sự liên kết các khái niệm.

- *Mức độ 3* (Đa cấu trúc): Người học xem xét nhiều yếu tố có liên quan nhưng chưa thể hợp nhất lại với nhau, tức là các ý tưởng vẫn còn lỏng lẻo, có sự liên kết các khái niệm nhưng mục tiêu chưa sâu.

- *Mức độ 4* (Xác lập quan hệ): Người học thể hiện các kết nối đầy đủ, tích hợp các phần lại với nhau để tạo nên một cấu trúc hoàn chỉnh và có nghĩa. Ở mức này người học có thể giải thích nguyên nhân, phân tích mối quan hệ, so sánh và tương phản.

- *Mức độ 5* (Trừu tượng mở rộng): Người học có thể đi xa hơn vấn đề đặt ra, liên kết với nhiều khái niệm khác, khái quát hóa để nắm bắt những vấn đề mới và trừu tượng hơn, ngoài ra, người học có khả năng đánh giá, dự đoán, sáng tạo và phản ánh.

Dữ liệu định tính được thu thập thông qua phân tích bài làm của các nhóm đối với mỗi câu hỏi nhằm mục đích cung cấp cái nhìn tổng quan về suy luận thống kê của SV.

2.5. Kết quả nghiên cứu

Ở trong bài viết này, chúng tôi tập trung vào phân tích câu trả lời của các nhóm SV đối với câu hỏi số 2 trong vấn đề đặt ra (xem Hình 1). Để tìm mối quan hệ giữa số con trong gia đình của các thành viên trong lớp và độ tuổi của mẹ, các nhóm đã thực hiện quá trình tổ chức, biểu diễn, phân tích, giải thích dữ liệu khá đa dạng và được trình bày tóm tắt dưới đây:

- Nhóm 1 không tổ chức dữ liệu thành dạng ngắn gọn, giữ nguyên 57 thông tin, chỉ sắp xếp lại bảng dữ liệu theo gia đình có số con giảm dần (từ 6 con đến 1 con), tô 2 màu khác nhau để phân biệt 2 nhóm dữ liệu, nhóm từ 4 con trở lên và nhóm từ 1 đến 3 con, và không vẽ biểu đồ (xem Hình 2).

Vì vậy, nhóm đã không thể phân tích dữ liệu, chỉ đưa ra các nhận xét chung và thiếu cơ sở thuyết phục, chẳng hạn như số con đông thường là những gia đình có số tuổi mẹ cao, các mẹ có độ tuổi 50 trở xuống có 2 con chiếm đa số, các mẹ có độ tuổi từ 39 tuổi trở xuống chỉ có 2 con.

STT	TP / không TP	Tuổi mẹ	Số con
1	không TP	78	6
2	không TP	55	5
3	không TP	55	5
4	không TP	56	5
5	không TP	58	5
6	Thành phố	62	5
7	Thành phố	66	5
8	không TP	70	5
9	không TP	80	5
10	Thành phố	84	5
11	không TP	55	4
12	Thành phố	56	4
13	không TP	60	4
14	không TP	61	4
15	Thành phố	65	4
16	không TP	65	4
17	không TP	67	4
18	Thành phố	77	4
19	Thành phố	82	4
20	Thành phố	82	4
21	Thành phố	84	4
22	Thành phố	35	3
23	Thành phố	38	3
24	không TP	45	3
25	Thành phố	52	3
26	Thành phố	55	3
27	không TP	56	3
28	Thành phố	57	3
29	không TP	60	3
30	Thành phố	60	3
31	không TP	60	3
32	Thành phố	60	3
33	Thành phố	60	3
34	không TP	61	3
35	không TP	62	3
36	Thành phố	68	3
37	Thành phố	78	3
38	không TP	93	3
39	Thành phố	35	2
40	Thành phố	38	2
41	không TP	38	2
42	Thành phố	38	2
43	không TP	39	2
44	Thành phố	42	2
45	Thành phố	43	2
46	không TP	50	2
47	Thành phố	50	2
48	Thành phố	51	2
49	Thành phố	54	2
50	Thành phố	54	2
51	Thành phố	55	2
52	Thành phố	55	2
53	không TP	58	2
54	không TP	59	2
55	không TP	60	2
56	Thành phố	64	2
57	Thành phố	68	1

Hình 2: Bảng thống kê của nhóm 1

- Nhóm 2 lập bảng thống kê số con trung bình theo nhóm tuổi của mẹ (35; 45], ..., (75; 95] và vẽ biểu đồ cột đơn để biểu diễn số liệu từ bảng, sau đó dựa vào biểu đồ để so sánh số con trung bình giữa các nhóm tuổi và rút ra kết luận là những gia đình có tuổi mẹ càng lớn thì số con càng đông. Ngoài ra, nhóm còn giải thích sự biến thiên của dữ liệu dựa vào bối cảnh mà vấn đề được đặt ra, đó là thời điểm tiếp cận chủ trương kế hoạch hoá gia đình, cách đây 20 năm, của các nhóm tuổi. Nghĩa là nếu tính lui 20 năm thì nhóm tuổi từ 65 đến 95 đã qua độ tuổi sinh con nên không bị ảnh hưởng bởi chủ trương này. Vì vậy, tỉ lệ sinh của nhóm tuổi này cao hơn, đối với các nhóm tuổi nhỏ hơn 65 thì sẽ có tỉ lệ sinh giảm dần.

- Nhóm 3 lập bảng thống kê số con trung bình và phương sai theo nhóm tuổi của mẹ: (35; 45] (75; 95], nhưng không vẽ biểu đồ (xem Hình 3). Đây là

nhóm đã khai thác dữ liệu ở nhiều khía cạnh khác nhau như so sánh số con trung bình theo nhóm tuổi của mẹ với số con trung bình của lớp, nhận xét về nhóm tuổi mẹ có số con trung bình thấp nhất, cao nhất, nhóm tuổi có mức độ chênh lệch số con nhiều nhất, ít nhất.

Trung bình số con	Lớp tuổi mẹ	Trung bình số con theo độ tuổi	Phương sai
3,23	[35;45]	2,22	0,19
	[45; 55]	2,29	0,24
	[55; 65]	3,32	1,16
	[65; 75]	4,00	0,00
	[75; 95]	4,11	0,36

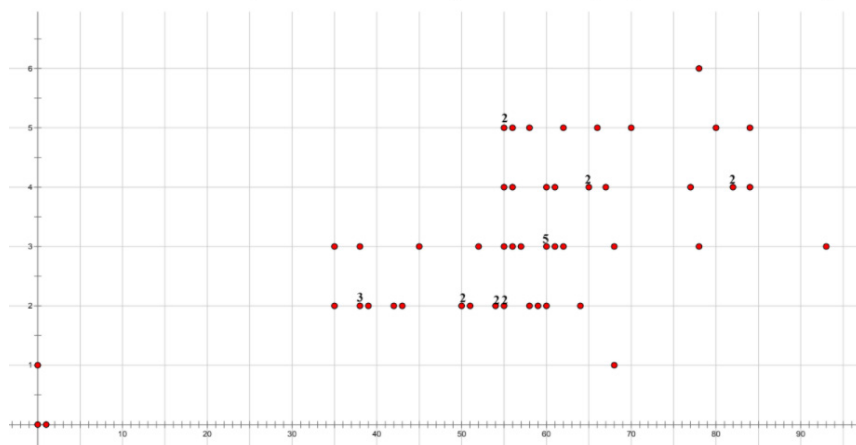
Từ đó ta thấy rằng:

- Trung bình số con của mỗi gia đình là 3,23.
- Mẹ trong độ tuổi [35; 45) có số con trung bình thấp nhất: 2,22.
- Mẹ trong độ tuổi [75; 95] có số con trung bình cao nhất: 4,11.
- Mẹ trong độ tuổi [55; 65) có mức độ chênh lệch số con lớn nhất, (phương sai là 1,16).
- Mẹ trong độ tuổi [65; 75) có mức độ chênh lệch số con nhỏ nhất, (phương sai là 0).

Hình 3: Bài làm của nhóm 3

Kết luận rằng, nếu độ tuổi của mẹ càng lớn thì số con trung bình càng cao. Số con trung bình của các mẹ có độ tuổi từ 65 trở lên lớn hơn số con trung bình của các gia đình trong lớp và ngược lại số con trung bình của các mẹ có độ tuổi từ 65 trở xuống thì thấp hơn, mức độ chênh lệch về số con giữa các gia đình cũng khác nhau theo độ tuổi của mẹ.

- Nhóm 4, đầu tiên, sắp xếp lại bảng dữ liệu theo 2 cách là tuổi mẹ tăng dần và số con giảm dần, dựa vào bảng số liệu SV chỉ rút ra được nhận xét chung chung như nhóm 1 đó là số gia đình có 3 con trở lên thì phần lớn tuổi mẹ trên 50, gia đình có 4 con trở lên thì tuổi mẹ đều trên 55. Tiếp theo, để có một cách nhìn sâu hơn về dữ liệu, nhóm này đã vẽ một biểu đồ để biểu thị mối tương quan giữa 2 biến là tuổi mẹ và số con, nhưng SV không biết đó là biểu đồ phân tán. Thông qua quan sát xu hướng tập trung của dữ liệu thể hiện trên biểu đồ, SV đã nhóm tuổi của mẹ theo các lớp (35;45],..., (75;95] rồi tính số con trung bình theo nhóm tuổi của mẹ và so sánh giữa các nhóm. Mặc dù không trình bày bảng thống kê dưới dạng ngắn gọn, nhưng các em đã tổ chức lại dữ liệu thông qua việc tính số con trung bình theo nhóm tuổi của mẹ, từ đó có cơ sở để so sánh, phân tích và rút ra kết luận (xem Hình 4).

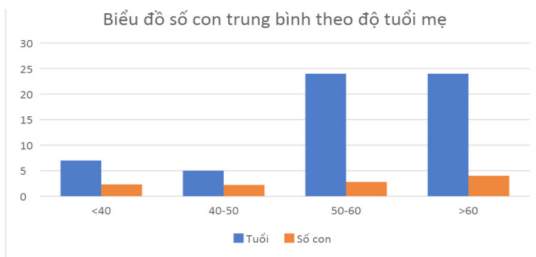
Biểu đồ thể hiện số con của các gia đình trong lớp theo tuổi của mẹ:**Hình 4: Biểu đồ của nhóm 4**

- Nhóm 5 lập bảng thống kê số con trung bình theo nhóm tuổi của mẹ là (30;40]..., (90;100]. Nhóm này không vẽ biểu đồ mà dựa vào bảng thống kê để so sánh, nhận xét số con trung bình ở các nhóm tuổi, sau đó dựa vào bối cảnh xã hội để giải thích kết quả. Chẳng hạn như nhóm tuổi của mẹ từ 70 đến 80 và từ 80 đến 90 có tỉ lệ sinh cao nhất vì trong độ tuổi sinh của các mẹ thì chưa có chính sách kế hoạch hóa gia đình; các mẹ trong nhóm tuổi 30 đến 50 có tỉ lệ sinh thấp nhất vì trong độ tuổi sinh của các mẹ Nhà nước đã áp dụng chính sách “mỗi gia đình có từ một đến hai con” nên tỉ lệ sinh giảm hẳn (2,28 đến 2,33) nhưng vẫn cao hơn tỉ lệ sinh chung của cả nước.

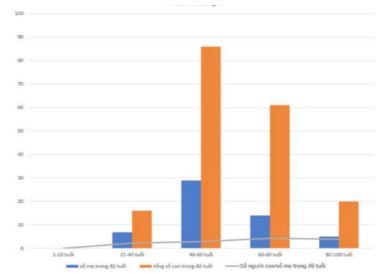
- Nhóm 6 không tóm tắt dữ liệu và cũng không vẽ biểu đồ mà dựa vào bảng số liệu ban đầu để đưa ra nhận xét chung chung, thiếu cơ sở, lập luận mang tính chủ quan. Cụ thể nhóm này đã đưa ra câu trả lời như sau: “Tuổi mẹ càng cao thì nhìn chung số con càng nhiều. Vì lúc này người mẹ đã hoàn thành việc sinh con. Ngược lại với những người mẹ còn trẻ thì vẫn có khả năng sinh thêm con. Độ tuổi dưới 40 thường sinh 2 con theo chủ trương kế hoạch hóa gia đình. Độ tuổi trên 50 thường sinh 3 đến 4 con vì họ thuộc thế hệ trước chưa quy định ko sinh con thứ 3”.

- Nhóm 7 lập bảng 2 chiều về số con và nhóm tuổi của mẹ (dưới 40, 40 đến 50, 50 đến 60 và trên 60) (xem Hình 5). Sau đó, SV tiếp tục tính số con

trung bình ứng với mỗi nhóm tuổi và vẽ biểu đồ tần số cột kép về độ tuổi mẹ và số con trung bình. Dựa vào biểu đồ, nhóm so sánh số con trung bình ở các nhóm tuổi của mẹ, và rút ra kết luận độ tuổi mẹ tỉ lệ thuận với số con trong gia đình, và tỉ lệ đông con chủ yếu tập trung ở các gia đình có tuổi của mẹ từ 50 trở lên. Mặc dù nhóm vẽ biểu đồ cột kép nhưng biểu đồ tần số về độ tuổi của mẹ (màu xanh) không được sử dụng để phân tích, giải thích.



Hình 5: Biểu đồ của nhóm 7



Hình 6: Biểu đồ của nhóm 8

- Nhóm 8 lập bảng thống kê số con trung bình theo nhóm tuổi của mẹ: (20;40], (80;100]. Vẽ biểu đồ kết hợp gồm biểu đồ tần số cột kép về độ tuổi của mẹ và tổng số con, biểu đồ đường biểu thị số con trung bình theo độ tuổi của mẹ (xem Hình 6). Tuy nhiên, biểu đồ tần số về độ tuổi của mẹ và tổng số con tương ứng không được SV sử dụng trong phân tích, giải thích, nhận xét mà các em chỉ dựa vào biểu đồ đường để đưa ra kết luận độ tuổi của mẹ càng lớn thì số con càng nhiều. Nhóm cũng dựa vào bối cảnh xã hội để giải thích kết quả tương tự như nhóm 5.

- Nhóm 9 không lập bảng thống kê, nhóm dữ liệu theo số con trong mỗi gia đình (1 con, 2 con, 6 con) và tính tuổi trung bình của mẹ tương ứng với các nhóm, vẽ biểu đồ cột kép gồm số con và tuổi trung bình của mẹ, các em đã có sự kết nối giữa 2 biểu đồ để giải thích cho kết luận đưa ra. SV dựa vào biểu đồ để nhận xét rằng đa số những gia đình có đông con (4 - 6 con) thì độ tuổi của mẹ cao hơn độ tuổi của mẹ ở những gia đình ít con (1 - 3 con). Theo nhóm này, có nhiều nguyên nhân dẫn đến kết quả trên, tức là tỉ lệ sinh ở các gia đình ngày càng giảm, đó là do chính sách kế hoạch hóa gia đình được tuyên truyền rộng rãi, điều kiện kinh tế xã hội, tâm lý ngại sinh, nhiều cặp vợ chồng hiếm muộn con... Như vậy, từ việc phân tích bài làm của SV, chúng tôi nhận thấy các biểu hiện liên quan đến suy luận thống kê như sau:

Khả năng tổ chức, biểu diễn và tóm tắt dữ liệu: Phần lớn các nhóm đều có khuynh hướng thu gọn bảng số liệu vì có khá nhiều giá trị khác nhau, bằng cách phân chia dữ liệu theo các lớp. Có 6/9 nhóm đã tổ chức lại dữ liệu bằng cách nhóm tuổi của mẹ thành các lớp với độ rộng của mỗi lớp là 10 hoặc 20. Ngoại trừ nhóm 4 đã dựa vào biểu đồ phân tán để nhận ra xu hướng tập trung của dữ liệu từ đó chia nhóm tuổi của mẹ, các nhóm còn lại không đưa ra cơ sở của việc phân lớp, điều này có thể giải thích là do các bài tập trong sách giáo khoa đều yêu cầu phân chia dữ liệu theo lớp được cho sẵn nên SV nghĩ rằng, có thể chia thành các lớp tùy ý miễn là thuận lợi cho việc thu gọn. Có 1 nhóm (nhóm 9) phân lớp số liệu thành 6 lớp theo số con trong mỗi gia đình. Có 2/9 nhóm không tổ chức dữ liệu thành dạng ngắn gọn (bảng hoặc biểu đồ) nên không thấy được mô hình hoặc xu hướng của dữ liệu, vì vậy chỉ đưa ra những nhận định chung chung, thiếu suy luận, không có cơ sở thuyết phục. Về biểu đồ, có 4/9 nhóm không vẽ, trong đó 2 nhóm (nhóm 3 và nhóm 5) nhận thấy rằng, dựa vào bảng thống kê thu gọn đã có thể phân tích và so sánh nên biểu đồ là không cần thiết. 5/9 nhóm đã sử dụng các loại biểu đồ khác nhau như biểu đồ cột, biểu đồ cột kép, biểu đồ kết hợp cột kép và đường gấp khúc, biểu đồ phân tán để thuận lợi cho việc quan sát cấu trúc của dữ liệu. Tuy nhiên, đối với nhóm 2 và nhóm 8, việc vẽ thêm biểu đồ không phát huy tác dụng nào hơn so với bảng thống kê mà nhóm đã lập trước đó. Trong 9 nhóm, chỉ có 1 nhóm sử dụng biểu đồ phân tán mô tả mối tương quan của hai đối tượng là tuổi mẹ và số con, từ đó bước đầu nhận ra xu hướng của dữ liệu thông qua quan sát, nhưng các em chưa biết cách tính các thông số khác như hệ số tương quan, đường xu hướng để có thể đưa ra những kết luận chính xác hơn. Nhóm 7, 8 và 9 vẽ các biểu đồ cột kép và biểu đồ kết hợp, nhưng chỉ có nhóm 9 biết kết hợp 2 biểu đồ để giải thích kết luận của mình, nhóm 7 và 8 chỉ sử dụng 1 biểu đồ trong số đó vào bước phân tích. Như vậy, SV chỉ vẽ biểu đồ vì có thể biểu diễn dữ liệu một cách trực quan mà chưa quan tâm đến tính hiệu quả cũng như lựa chọn biểu đồ phù hợp với đặc điểm của mẫu dữ liệu. Để vẽ biểu đồ, phần lớn các nhóm sử dụng công cụ Chart của phần mềm Word hoặc Excel, một nhóm tự vẽ biểu đồ phân tán bằng phần mềm Sketchpad.

Quá trình phân tích và giải thích dữ liệu, hầu hết các nhóm đều nhận ra xu hướng của dữ liệu là những gia đình có tuổi mẹ càng lớn thì số con càng đông và ngược lại. 7/9 nhóm dùng số trung bình kết hợp với so sánh dữ liệu để

nhận xét về mối quan hệ giữa số con trong gia đình của các thành viên trong lớp và độ tuổi của mẹ, chỉ có 1 nhóm sử dụng thêm phương sai để đo mức độ phân tán của dữ liệu. Bên cạnh đó, có 4 nhóm đã khai thác bối cảnh mà vấn đề đặt ra để giải thích tính hợp lí của kết quả thu được. Không có nhóm nào quan tâm đến độ rộng của dữ liệu, tính hợp lí của các số liệu chẳng hạn như tuổi mẹ nhỏ nhất là 35 và lớn nhất là 93, trong khi đó tuổi của SV trong lớp nhỏ nhất là 21 và lớn nhất là 23. Ngoài ra, theo mô tả mức độ của phân loại SOLO và phân tích bài làm của các nhóm, chúng tôi có được kết quả là 6/9 nhóm thuộc mức độ 4 (gồm các nhóm 2, 3, 4, 5, 8 và 9), 1/9 nhóm thuộc mức độ 3 (nhóm 7) và 2/9 nhóm thuộc mức độ 2 (nhóm 1 và 6). Như vậy, hai phần ba số nhóm đã xem xét nhiều hơn một yếu tố trong quá trình suy luận, đồng thời liên kết các yếu tố có liên quan với nhau, phân tích mối quan hệ, so sánh để đưa ra kết luận, giải thích kết quả theo bối cảnh của vấn đề.

3. KẾT LUẬN

Kết quả nghiên cứu cho thấy, mặc dù hai phần ba số nhóm tham gia thực nghiệm thuộc mức độ 4/5 của thang đo SOLO nhưng quá trình phân tích định tính bộc lộ những khiếm khuyết trong suy luận thống kê của SV cũng như chương trình giảng dạy Thống kê ở bậc Phổ thông hiện tại. Nghiên cứu cũng chứng tỏ, việc sử dụng quy trình khảo sát thống kê tạo điều kiện cho SV thể hiện suy luận thống kê và đem lại hứng thú học tập cho các em, giúp các em hiểu rõ hơn ý nghĩa của Thống kê trong thực tế, góp phần phát triển khả năng vận dụng kiến thức Thống kê vào thực tiễn. Trên cơ sở của nghiên cứu, chúng tôi có một số đề xuất đối việc dạy học Thống kê giúp nâng cao năng lực thống kê của người học mà SV Sư phạm ngành Toán cần được rèn luyện trong quá trình học tập ở trường ĐH. Đó là, trong lớp học, GV nên chú trọng vào những ý tưởng để giúp HS hiểu khái niệm và các mối quan hệ giữa những khái niệm hơn là tập trung vào các công cụ tính toán. Bên cạnh đó, GV cần tạo điều kiện để HS thực hành các phương pháp thu thập và tinh giản dữ liệu vì điều này ảnh hưởng đến chất lượng của dữ liệu cũng như các kiểu phân tích thích hợp, đồng thời HS nên được lựa chọn các mô hình trực quan để mô tả dữ liệu một cách phù hợp, hiệu quả trong từng tình huống. GV nên sử dụng các tập dữ liệu thực để tạo sự quan tâm chú ý cho HS cũng như khuyến khích các em

đưa ra và kiểm tra các phỏng đoán, sử dụng các hoạt động trong lớp học được thiết kế cẩn thận để hỗ trợ và nâng cao khả năng suy luận của HS, đẩy mạnh thuyết trình trong lớp học, bao gồm các lập luận thống kê và tăng cường trao đổi những ý tưởng thống kê quan trọng. Ngoài ra, GV có thể khuyến khích HS sử dụng các công cụ công nghệ thích hợp để kiểm tra những phỏng đoán của mình và phân tích dữ liệu, từ đó phát triển khả năng suy luận thống kê. Vì Thống kê gắn liền với thực tiễn, GV có thể xem Thống kê là một trong những nội dung đem lại nhiều cơ hội để dạy học mô hình hóa, giúp cho HS hiểu được ý nghĩa của tri thức thống kê và rèn luyện được tư duy thống kê cho HS, nâng cao sự hiểu biết của các em về Thống kê.

TÀI LIỆU THAM KHẢO

- [1] Frost, J. (2013), Why statistics is important, *The World of Statistics*, Retrieved from <http://www.worldofstatistics.org/2013/03/04/why-statistics-is-important/>.
- [2] Pfannkuch, M., & Ben-Zvi, D. (2011), Developing teachers' statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education*, pp. 323-333, Springer, Dordrecht.
- [3] Garfield, J., & Ben-Zvi, D. (2008), Preparing school teachers to develop students' statistical reasoning, *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study*, 299-310.
- [4] Ben-Zvi, D., & Garfield, J. (2004), Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking*, pp. 3-16, Dordrecht, The Netherlands: Kluwer Academic Publishers.
- [5] Cobb, G. W., & Moore, D. (1997), *Mathematics, statistics, and teaching*, American Mathematical Monthly, 104, 801-823.
- [6] Rossman, A. J., Chance, B. L., & Medina, E. (2006), Some important comparisons between statistics and mathematics, and why teachers should care. In G. F. Burrill (Ed.), *Thinking and reasoning about data and chance:*

- Sixty-eighth NCTM Yearbook*, pp. 323–334, Reston, VA: National Council of Teachers of Mathematics.
- [7] Garfield, J., and Gal, I. (1999), Teaching and Assessing Statistical Reasoning, In *Developing Mathematical Reasoning in Grades K-12*, ed. L. Stiff, Reston, VA: National Council Teachers of Mathematics, 207-219.
- [8] Garfield, J. and Ben-Zvi, D. (2007), *How students learn statistics revisited: A current review of research on teaching and learning statistics*, International Statistical Review, 75(3), 372–396.
- [9] delMas, R. (2002), *Statistical literacy, reasoning, and thinking: A commentary*, Journal of Statistics Education, 10(3).
- [10] Vygotsky, L. (1978), *Mind in society*, Cambridge, MA: Harvard University Press.
- [11] Cobb, P. and McClain, K. (2004), Principles of instructional design for supporting the development of students’ statistical reasoning, In: D. Ben-Zvi and J. Garfield (eds) *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, pp. 375–396, Dordrecht, The Netherlands: Kluwer Academic Publishers.
- [12] Wild, C. J., & Pfannkuch, M. (1999), *Statistical thinking in empirical enquiry*, International Statistical Review, 67(3), 223-265.
- [13] Biggs, J. B., & Collis, K. F. (1982), *Evaluating the quality of learning: The SOLO taxonomy*, New York: Academic Press.
- [14] Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2012), Technology for enhancing statistical reasoning at the school level, In *Third international handbook of mathematics education*, pp. 643-689, Springer, New York, NY.
- [15] Chance, B. L. (2002), *Components of statistical thinking and implications for instruction and assessment*, Journal of Statistics Education, 10 (3).
- [16] DelMas, R. (2003), A comparison of mathematical and statistical reasoning, In D. BenZvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Reasoning, Literacy, and Thinking*. Dordrecht: Kluwer.
- [17] Garfield, J. (2002), *The challenge of developing statistical reasoning*, Journal of statistics education, 10 (3).

- [18] Garfield, J., & Ben-Zvi, D., (2009), *Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment*, Teaching Statistics, 31 (3), 72-77.

EXPLORING STUDENTS' STATISTICAL REASONING THROUGH STATISTICAL INVESTIGATION ACTIVITY

Abstract: *In this study, the authors focus on the statistical reasoning of Mathematics pre-service teachers through a statistical investigation. The research results are used to formulate recommendations to develop statistical reasoning for Mathematics students at universities of education. Experimental method is used to observe and analyze students' statistical reasoning. Although two-thirds of the experimental groups are at level 4/5 of the SOLO scale, the qualitative analysis reveals shortcomings of the students' statistical reasoning as well as the current high school statistics curriculum. The study also demonstrates that using statistical investigation process in teaching statistics creates opportunity for students to express statistical reasoning and bring excitement in learning, helping them better understand the meaning of statistics in real world and develop the ability to apply statistical knowledge in practice.*

Keywords: *Statistical investigation process, statistical reasoning, statistics.*