Students' Understanding of Test Statistics in Hypothesis Testing

Annie Burns-Childers
University of Arkansas Little Rock

Darryl Chamberlain Jr.
Georgia State University

Aubrey Kemp
Georgia State University

Leslie Meadows
Georgia State University

Harrison Stalvey
University of Colorado Boulder

Draga Vidakovic
Georgia State University

Hypothesis testing is a key concept included in many introductory statistics courses. Yet, due to common misunderstandings of both scientists and students, the use of hypothesis testing to interpret experimental data has received criticism. With statistics education on the rise as well as an increasing number of students enrolling in introductory statistics courses each year, there is a need for research that investigates students' understanding and curriculum effectiveness of hypothesis testing. This paper describes results obtained from a larger study designed to explore introductory statistics students' understanding of one sample hypothesis testing. In particular, this paper explores students' understanding of test statistic as a component of hypothesis testing. APOS Theory is used as a guiding theoretical framework. This paper focuses on three students' understandings of test statistic when performing hypothesis tests on real world data.

Key Words: Hypothesis Testing, Statistics, Test Statistic

Introduction

The use of statistics is crucial for numerous fields, such as business, medicine, education, and psychology. Due to its importance, according to the *Guidelines for Assessment and Instruction in Statistics Education College Report*, more students are studying statistics, and at an increasingly younger age (GAISE College Report ASA Revision Committee, 2016). In the United States today, the *Common Core State Standards for Mathematics* calls for students to "understand statistics as a process for making inferences about population parameters based on a random sample from that population" (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, p. 81). More recently, the *GAISE College Report* calls for nine goals for students in introductory statistics courses, including "Students should demonstrate an understanding of, and ability to use, basic ideas of statistical inference, both hypothesis tests and interval estimation, in a variety of settings" (GAISE College Report ASA Revision Committee, 2016, p. 8). In other words, in an introductory statistics course, students should understand and be able to apply hypothesis testing in various situations.

Hypothesis testing is an important tool of statistical inference (Krishnan & Idris, 2015). However, the use of hypothesis testing to interpret experimental data has received criticism (Nickerson, 2000; Nuzzo, 2014) due to the common misunderstandings of both scientists and students when using this method (Batanero, 2000; Dolor & Noll, 2015; Vallecillos, 2000). Rather than abandon the inference method entirely, researchers have called for improving the education and understanding of hypothesis testing. For example, LeMire (2010) developed a framework to revise and improve instructional content on hypothesis testing. Even still, there are few studies on student understanding of hypothesis testing as a whole (Smith, 2008). Our larger research goal is to supplement these efforts by first analyzing how students come to understand hypothesis testing and then develop instructional materials to cultivate this understanding. What follows is a brief summary of the literature on student understanding of hypothesis testing.

Literature Review

Research suggests that although students are able to perform the procedures surrounding hypothesis testing, students lack a strong understanding of the concepts and their use (Smith, 2008). It is suggested that hypothesis testing and the area of inferences "is probably the most misunderstood, confused and abused of all statistical topics" (Batanero et al., 1994, p. 541). Textbooks and instructors often give a specific step-by-step script to follow when performing hypothesis testing, not connecting or allowing the students to see the idea as a whole. Link (2002) suggested this practice as a six-part procedure, which leads many students to look for key words and phrases as guides when solving hypothesis testing problems. He also found evidence that supports the belief that students can correctly substitute values into a formula selected from a formula sheet, yet they do not have a full understanding of hypothesis testing in its entirety.

In an attempt to move away from a procedural approach, and due to the rise of statistical education, calls for reform have led to a shift from an emphasis on procedural understanding to conceptual understanding (GAISE College Report ASA Revision Committee, 2016; Krishnan & Idris, 2015). Ways to teach for conceptual understanding have been varied. For example, Hong & O'Neil (1992) suggested that to foster this conceptual understanding in hypothesis testing, conceptual instruction should be presented prior to the procedural instruction with emphasis on the use of diagrammatic problem representations. In contrast, Chandrakantha (2014) suggested that utilizing technology that allows students to visualize and work hands-on with data will enhance student understanding of concepts such as hypothesis testing.

Other research focuses on identifying students' misconceptions with various parts of hypothesis testing in order to improve conceptual understanding. Specifically in introductory statistics courses, students appear to experience a "symbol shock" (Schuyten, 1990), which provides an obstacle for students interpreting particular questions (Dolor & Noll, 2015; Lui & Thompson, 2005; Vallecillos, 2000). Vallecillos (2000) found that students have trouble with not only the symbols, but also with the formal language and meaning behind the concepts involved in hypothesis testing, including words such as 'null' and 'alternative' hypothesis. Students interviewed were not able to accurately describe what these terms mean and how they impact the decision to either accept or reject within the test (Vallecillos, 2000). Furthermore, Williams (1997) during two different interviews with 18 students, found that the term 'significance' was not well understood. Students gave vague and inadequate descriptions of what 'significance' means in the context of hypothesis testing. Even after the final exam, students continued to confuse terminology and still had a poor understanding of the concept. Mathematical symbolism presents challenges across all levels (Rubenstein, 2008), especially as most students' school mathematics experiences give little or no attention to the idea of reading mathematics as a language (Adams, 2003).

With statistics education reform on the rise, as well as an increasing number of students enrolling in introductory statistics courses each year, there is a need for research that investigates students' understanding and curriculum effectiveness of hypothesis testing, a concept taught in almost every introductory statistics course (GAISE College Report ASA Revision Committee, 2016; Krishnan & Idris, 2015). Literature in this area focuses on students' lack of understanding of hypothesis testing, ways to teach for conceptual understanding of hypothesis testing, and misconceptions with various parts of hypothesis testing. Particular parts of hypothesis testing that have received the most attention in literature are hypotheses and *p*-value. Test statistics contribute to understanding a *p*-value, but there is a lack of literature on this concept (Smith, 2008). Thus, we focus our attention on the following research question:

> What are students' understandings of test statistic, as a component of hypothesis testing as a whole, in two distinguished real world situations?

The next section will introduce test statistics within the scope of one-proportion hypothesis testing.

Test Statistic

A test statistic, according to the textbook *Elementary Statistics Using Excel*, is "a value used in making decisions about the null hypothesis," (Triola, 2014, p. 415). While the definition provided is simplistic, the actual concept of test statistic in hypothesis testing is complex. Assuming the null hypothesis is true, a test statistic is found by converting a sample statistic, such as a sample proportion or a sample mean, to a standardized score. Although we discuss test statistic as one mathematical term or value, there is a distinction between test statistics calculated from sample proportions versus sample means. Specifically, in our study, the distinction is between the normal distribution and the Student's *t* distribution.

When calculating a test statistic for a sample representing a population mean or proportion, we are referring to a standardized value that represents the extremeness of a sample in regards to what is expected. For proportions, students use the normal distribution as an approximation to the binomial distribution, and thus calculate test statistics, which are *z*-scores. For means, students learn about test statistics in hypothesis testing using the normal distribution (*z*-scores) and the Student's *t* distribution (*t*-scores). Although these distributions appear similar, the distinction occurs depending on what we know about our sample. In particular, if we know the population standard deviation, then we know the center, shape, and spread of the data, and can use the normal distribution (*z*-scores). However, if the population standard deviation is unknown and we only have the sample standard deviation as an estimate, then we can use the Student's *t* distribution to represent the center, shape and spread of the data. Because of the greater variability associated with smaller sample sizes, *t*-scores are greater than or equal to *z*-scores for the same value of n, but as n approaches infinity, the limiting value of the Student's t distribution is the normal distribution (see Figure 1).

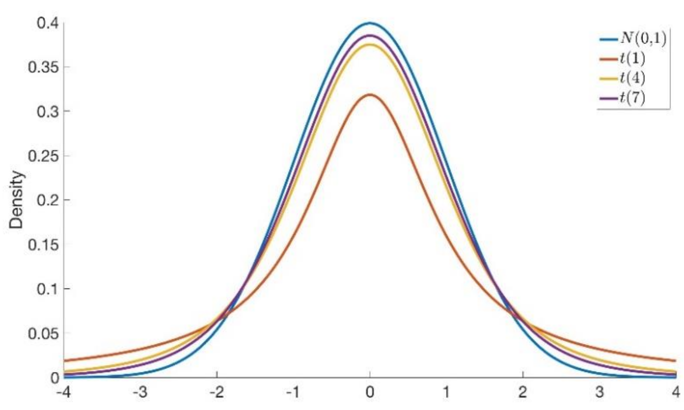

Figure 1. Graphical Representation of Normal Distribution and Student's t Distribution
(Note: key represents $t(n)$, where n is the sample size)

Theoretical Framework

A theoretical framework is necessary in order to analyze and describe how students understand a particular concept. The guiding theoretical framework for our larger study is APOS Theory (Asiala et al., 1996). APOS Theory is a framework which models an individual's mathematical conception using Actions, Processes, Objects, and Schema. An Action is an externally driven transformation of a mathematical object (or objects). An Action can be described as an individual needing an external cue to complete a task, such as a step-by-step example to follow. For our study, in the context of test statistic, an example of this would be a student calculates the test statistic by the correct formula (even with wrong distribution label), but cannot interpret the test statistic verbally or graphically. Another example, in the context of test statistic, would be a student who uses key words or phrases from the problem as external cues to decide which formula to use to calculate the test statistic.

Once Actions are repeated and reflected on, an individual can start to interiorize them to become a Process. A Process no longer requires step-by-step external cues. An individual is now able to internally imagine the steps in a transformation, without having to actually perform them for specific examples. In the context of test statistic, an example would be a student calculates test statistic by using the correct formula (even with wrong distribution label) and describes their calculations in general terms. Although the student can calculate the test statistic, they do not have to perform the steps in order to describe the calculations. Another example would be a student who interprets the test statistic verbally or graphically by describing how it corresponds to $\hat{p}$ or $\bar{x}$ or the distance from the sample value to the expected value ($\hat{p}$ to p or $\bar{x}$ to μ).

When an individual is then able to see the Process as a totality, is aware that transformations can be applied to it, then the Process has been encapsulated into an Object. In the context of test statistic, several examples of this are as follows: a student describes the test statistic as an input of a function that determines the p-value; student interprets test statistic graphically as determining the edge of the region whose area is the p-value; or student describes the test statistic as large or small in comparison to a "usual" value of a test statistic (not just describing the number as large or small in general). As defined by the textbook, a "usual" value of a test statistic refers to z-scores between -2 and 2 (Triola, 2014).

Schemas are "structures that contain the descriptions, organization, and exemplifications of the mental structures that an individual has constructed regarding a mathematical concept" (Arnon et al., 2014, p. 25). Schemas may also be included within another Schema. For example, the distribution schema plays a role in the development of the schema of test statistic. This report is devoted to providing examples of Action and Process conceptions of test statistic from a sample of student interviews conducted for a larger study.

Methodology

The focus of our study is on university students who are enrolled in an introductory statistics course based on the emporium model. The emporium model, originated at Virginia Tech, includes key components of "interactive computer software, personalized on-demand assistance, and mandatory student participation" (Twig, 2011, p. 26). For this particular institution, each week students were required to spend three academic hours in a computerized mathematics lab, as well as attend one academic hour class each week with an instructor. The time in the mathematics lab was spent actively learning using the mathematical software *MyStatLab* by Pearson. Students also read, viewed videos, and discussed material with peers, lab assistants, and instructors.

Relevant data was collected during Fall 2014 and Spring 2015. All students enrolled in six sections of an introductory statistics course (approximately 240 students) were invited to participate in a problem solving session and semi-structured interview pertaining to hypothesis testing. Twelve students volunteered to participate. During the problem solving sessions, each participant worked alone on two hypothesis test questions. They were allowed to use Excel when needed since the use of it was required as part of the class. The first question asked the student to conduct and interpret a hypothesis test for a single population proportion. The second question asked the student to conduct and interpret a hypothesis test for a single population mean. The questions were as follows:

1. In a recent poll of 750 randomly selected adults, 588 said that it is morally wrong to not report all income on tax returns. Use a 0.05 significance level to test the claim that 70% of adults say that it is morally wrong to not report all income on tax returns. Use the *P*-value method. Use the normal distribution as an approximation of the binomial distribution.
2. Assume that a simple random sample has been selected from a normally distributed population and test the given claim. In a manual on how to have a number one song, it is stated that a song must be no longer than 210 seconds. A simple random sample of 40 current hit songs results in a mean length of 231.8 seconds and a standard deviation of 53.5 seconds. Use a 0.05 significance level to test the claim that the sample is from a population of songs with a mean greater than 210 seconds.

Students had seen these exact questions with variable parameters on their homework and quizzes when using the *MyStatLab* software. Students also engaged in active learning associated with these concepts both in the lab and in class. Thus, students were expected to know how to conduct and interpret hypothesis tests for both questions, and in particular, they were expected to know to use the normal distribution to find the test statistic for Question 1 and the Student's t distribution to find the test statistic for Question 2. In other words, they were expected to know the procedure.

Immediately following the problem solving sessions, the students participated in semi-structured interviews. There were ten interviews, eight with one participant each and two with two participants each (12 students in total). During the semi-structured interviews, participants were asked to elaborate on their answers and thought processes. The interviews were conducted and divided among multiple members of the research team. To standardize the interviews, an interview protocol was developed beforehand. The relevant data for this paper consists of participants' written work, Microsoft Excel files, and transcribed discussion from the follow up interviews.

Data analysis took place after all interviews were conducted. The recordings of the interviews were distributed and transcribed by each of the six members of the research team. After transcriptions were completed, analysis was organized in a way so that each transcript was reviewed by two different pairs of researchers. The data was coded and analyzed according to APOS Theory. After codes were developed and agreed upon by each pair of researchers, the team came together for discussion as a whole. Six concepts of hypothesis testing emerged in participants' reasoning. This paper focuses on one of these six concepts, test statistic. The data and codes were then used to develop individual learning trajectories for each participant that merely served as a method to explore and organize the ways of understanding of each concept for this group of individuals. The focus of this paper will be on the individual learning trajectories for test statistic.

Results

In this section, we illustrate examples from the data analysis that are most relevant and indicative of the Action and Process conceptions of test statistic according to APOS Theory. Our results will focus on three students: Haley, Lana, and Steve. Each student was representative of different subgroups of the twelve learning trajectories, and exhibited different conceptions of test statistic.

Haley

Haley demonstrated evidence of an Action conception of test statistic. For both questions, Haley looked for indicator phrases to identify which test statistic formula to use. She identified that a problem about proportions implies that the test statistic will be a *z*-score. Haley also mentioned that if in a problem she is not given the population standard deviation, this indicates when she would use a *t*-score.

I: Okay, and um is this a t-score or a z-score?
H: This is a z… yeah this is a z-score.
I: And how did you know that?
H: Because we're using proportions so you just use z.
I: Okay, good. Um, and then when would you use a t-score?
H: When you don't know the.. like when you don't know the um… what is it? You don't know… when you don't know this [draws sigma on the paper].
I: Okay!
H: But like you know it for a sample! [laughs]
I: So this is.. so.. I know that this is referring to the standard deviation right?
H: Yeah.
I: So and you said if you know the sample… so I'm assuming you're saying this is not the sample standard deviation?
H: No, that's the population. So if you don't know the population standard deviation I should use t.

Haley appeared to base her ideas of which test statistic to use off of key words found in the problems. When calculating the test statistic, Haley again looked for indicator phrases in the problem to identify the values to plug in. She also goes on later in the interview to state that she has memorized the formulas for test statistics.

I: Okay, then for the test statistic… how did you figure out all those numbers and stuff?
H: Well, I took the [yawn] I took x bar minus mu and then I divided it by the standard deviation divided by the square root of n.
I: Okay and did you just find that formula on the sheet or did you have that one memorized?
H: I had it memorized.

As more evidence of a lack of full understanding of what a test statistic represents, for the second problem, she identified the given sample mean in addition to her test statistic in different places on her graph when in actuality the test statistic is representative of the sample mean.

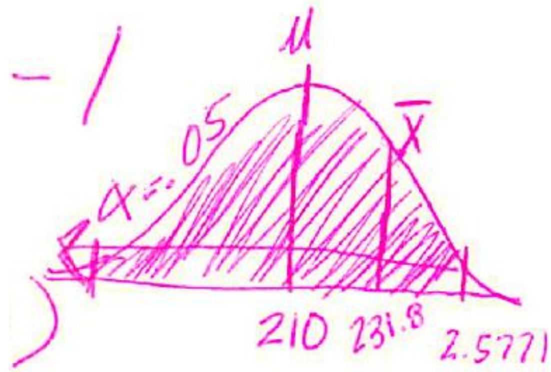

Figure 2. Haley's Graphical Representation of Question 2

Her memorization of formulas and what appears to be dependency on key words in the problem to identify which test statistic to use and what values to plug in, with no verbal explanation of the reasoning behind this, suggests Haley illustrates an Action conception of test statistic.

Lana

Lana exhibited evidence of a Process conception of test statistic. Lana, for Question 1, used the normal distribution to find the test statistic based on the fact that the problem was about proportions. After recognizing which test statistic to use, she described in general how to calculate the test statistic using the formula and she double checked all her work in Excel.

> L: And then I double checked by using the formula, the phat minus p over the square root of pq over n. I put that in Excel to double check and make sure that was right.

Lana was then prompted to explain the test statistic. She described, in words, what she imagined.

> L: I think that, I'm picturing the big curve, the bell curve, and I'm picturing the test statistic is where the point that falls on there … Okay, so this is the mean right in the middle, and the test statistic is one side of it, saying this is how far away from what they are saying is the mean, this is what the mean of this.

By reflecting on her previous steps and calculations, Lana appeared to interiorize the action of calculating test statistics. She internally imagined calculating or finding test statistics in relation to the mean, without having to actually perform any calculations. This suggests Lana exhibited a Process conception of test statistic. After her explanation, she drew a picture to illustrate her thinking (Figure 2) of a graphical representation of the normal distribution.

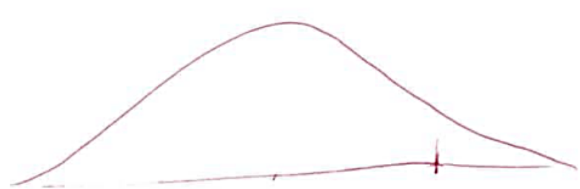

Figure 3. Lana's Graphical Representation of the Normal Distribution

In her written work, Lana initially used the normal distribution to find the test statistic for Question 2. During the interview she became worried when asked if the question was a *z*-test. She mentioned that she remembered from high school to use a *t*-score if the sample size is 30 or less. After prompting, she realized she should have used a *t*-score, however, she also immediately recognized that her *z*-score would "probably not" be very different from the *t*-score. She made the observation that *t*-scores and *z*-scores are "really close together". This is evidence of the role that distribution Schema plays in the development of the Schema of test statistic.

It appears that Lana illustrates a Process conception of test statistic because she pictured in her head a graphical interpretation of test statistic without relying on specific calculations, and also interpreted the test statistic as describing how far away the sample statistic is from the expected (population) value. It also appears that she possesses a distribution Schema that is in an early stage, as noted in her confusion between *z*-scores and *t*-scores. However, her understanding that a *z*-score and a *t*-score are "really close together" suggests that her distribution Schema is progressing and emerging.

Steve

Steve is included in the results as illustrative of a student whose analysis suggests discrepancies in his conceptual understanding of test statistic. For Question 1, Steve identified key words in the problem to recognize which test statistic to use. He explained, "I immediately thought of these two formulas, and at first I wasn't sure which one to use, and then I was like, oh wait, there's no x-bar or mu or standard deviation. So that makes it pretty easy." He used a system of elimination to decide which formula not to use. Even though he correctly identified the test statistic, he went on to say that this is "just a formula that I've learned like any other" and that he "doesn't understand why we use that formula, other than we just use it". He concluded his explanation stating that Question 1 used a *z*-value because the problem was of proportions, what appears is likely a memorized rule applied to the problem.

For Question 2, Steve mentioned that "it's basically the same problem", other than now being a question of means. Steve identified the distribution as normal based on the language in the problem, "a simple random sample has been selected from a 'normally distributed' population". He further explained that he would not know how to deal with a non-normally distributed population, that he could not recall learning anything other than a normal distribution, and that he did not know much about a Student's *t* distribution. Ironically, later in the interview, when asked if his answer was a *z*-value or *t*-value, he responded that it is "a t-value because you're testing means".

The excerpts above are suggestive of Action conception of test statistic. Steve appeared to rely on external cues based on key words identified in the problem and what appeared to be memorized rules of when to use a particular formula. However, when prompted by the interviewer to explain what a test statistic is, Steve elaborated as follows, "And also your test statistic is very large. I'm not totally sure what a test stat is, but it reminds me of z-scores, and I remember when you have a z-score that gets above 3. It starts to get pretty, pretty crazy. So 5 is huge, which is also the reason that you're getting a bunch of zeros or very close to 1 [for the p-value]." Although Steve's interpretation of z-scores seems fairly advanced in comparison to other students, it is not clear from his description whether he considers a *z*-score to be an example of the concept of test statistic or a similar concept that is distinct from the test statistic. The former would be suggestive of a Process conception of test statistic, while the latter would be suggestive of an Action conception of test statistic. Since these discrepancies exist, we suggest that Steve is emerging from the Action conception to the Process conception of understanding of test statistic.

Discussion and Concluding Remarks

Our results provide evidence of three students' understanding of test statistic, one who appears to exhibit an Action conception of test statistic, one who appears to exhibit a Process conception of test statistic, and one who appears to be emerging from an Action conception to a Process conception of test statistic. Consistent with Link (2002), many students used a procedural approach to hypothesis testing which included students plugging in correct values for a formula. In reference to test statistic, Haley looked for indicator phrases and key words in the problem to decide which test statistic formula to use and to decide which values to plug in, yet she did not have a full conceptual understanding of hypothesis testing. According to APOS Theory, she is illustrative of an Action conception of test statistic.

Another student, Lana, reflected on her calculations and appeared to interiorize the action of calculating test statistics by describing the test statistic in relation to the mean. However, for Question 2, Lana initially used the normal distribution to find the test statistic, rather than the Student's *t* distribution. In our study, we did not initially consider the role that the stage of distribution Schema would have in the development of the concept of test statistic. However, our results suggest that an individual can possess a Process conception of test statistic, while possessing only an early stage of Schema for distribution. Further research is suggested to explore the extent to which an individual's distribution Schema has in the development of the concept of test statistic.

Lastly, discrepancies in data analysis suggest Steve is emerging from an Action conception of test statistic to a Process conception of test statistic. Excerpts are suggestive that he was relying on key words and memorized formulas to decide which test statistic to use, evidence of Action conception of test statistic. However, Steve did verbally elaborate about *z*-scores, indicative of a possible Process conception. Nevertheless, inconsistency exists in the fact that Steve did not equate z-scores to the test statistic and instead suggested they remind him of each other. If Steve had a Process conception of test statistic, we would have expected him to note that test statistics are standardized values, which in this case *z*-scores are an example representative of test statistic as a whole.

As hypothesis testing is a key concept included in the majority of introductory statistics courses (Krishnan & Idris, 2015), and is arguably one of the most misunderstood statistical topics (Bantanero, 1994), it is important to continue research that investigates students' understanding and curriculum effectiveness of hypothesis testing. This paper, specifically focused on test statistic, a component of hypothesis testing, revealed students' understanding in two distinguished real world situations. By analyzing data according to APOS Theory, we have illustrated examples of Action conception, Process conception, and an example of a student who is emerging from Action conception to Process conception of test statistic. This knowledge can then be used to describe how students may develop an understanding of test statistic – an important concept within hypothesis testing and one that has not received much attention within the research community.

References

Arnon, I., Cottrill, J., Dubinsky, E., Oktac, A., Fuentes, S.R., Trigueros, M., et al. (2014). *APOS theory: A framework for research and curriculum development in mathematics education.* New York, NY: Springer.

Adams, T.L. (2003). Reading mathematics: More than words can say. *International Reading Association, 56*(8), 786-795.

Asiala, M., Brown, A., DeVries, D. J., Dubinsky, E., Mathews, D., & Thomas, K. (1996). A framework for research and development in undergraduate mathematics education. In J. Kaput, E. Dubinsky, & A. H. Schoenfeld (Eds.), *Research in collegiate mathematics education II* (pp. 1-32). Providence: American Mathematical Society.

Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning, 2*(1-2), 75-98.

Batanero, C., Godino, J.D., Vallecillos, A., Green, D.R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology, 25*(4), 527-547.

Chandrakantha, L. (2014). Visualizing and understanding confidence intervals and hypothesis testing using Excel simulation. *Electronic Journal of Mathematics and Technology, 8*(3).

Dolor, J., & Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal, 14*(1), 60-89.

GAISE College Report. ASA Revision Committee. (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report.* http://www.amstat.org/education/gaise.

Hong, E. & O'Neil, H.F. (1992). Instructional strategies to help learners build relevant mental models in inferential statistics. *Journal of Educational Psychology, 82*(2), 150-159.

Krishnan, S. & Idris, N. (2015). An overview of students' learning problems in hypothesis testing. *Jurnal Pendidikan Malaysia, 40*(2), 193-196.

LeMire, S.D. (2010). An argument framework for the application of null hypothesis statistical testing in support of research. *Journal of Statistics Education, 18*(2).

Link, W. C. (2002). *An examination of student mistakes in setting up hypothesis testing problems.* Louisiana-Mississippi Section of the Mathematical Association of America.

Lui, Y. & Thompson, P.W. (2005). Teachers' understanding of hypothesis testing. In S. Wilson (Ed.), *Proceedings of the twenty-seventh annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education*. Roanoke, VA: Virginia Polytechnic Institute and State University.

Lui, Y. & Thompson, P.W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies: An International Journal, 4*, 126-138.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core standards for mathematics.* Washington, DC: Authors.

Nickerson, R.S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241-301.

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, *506*, 150-152.

Rubenstein, R.N. (2008). *Mathematical symbolization: Challenges across levels*. Mexico: Paper presented at the 11th International Congress on Mathematical Education. Available from http://tsg.icme11.org/document/get/853.

Schuyten, G. (1990). Statistical thinking in psychology and education. *ICOTS, 3*, 486-489.

Smith, T.M. (2008). *An investigation into student understanding of statistical hypothesis testing.* Doctoral Dissertation. University of Maryland.

Triola, Mario. (2014). *Elementary statistics using Excel* (5th ed.). Upper Saddle River, NJ: Pearson.
Twigg, C. (2011, May-June). The math emporium: A silver bullet for higher education. *Change Magazine*, 25-34.
Vallecillos, A. (2000). Understanding the logic of hypothesis testing amongst university students. *JMD*, *21*, 101-123.
Williams, A.M. (1997). Students' understanding of hypothesis testing: The case of the significance concept. *MERGA*, *20*, 585-591.