American Psychological Association

Scholarship of Teaching and Learning in Psychology

2332-2101/19/$12.00

2019, Vol. 5, No. 2, 162–168
http://dx.doi.org/10.1037/stl0000141

TEACHER-READY THEORY REVIEW

Introducing the New Statistics in the Classroom

Christopher J. Normile, Emily K. Bloesch, Christopher C. Davoli, and Kyle C. Scherr
Central Michigan University

The new statistics, which include estimations of effect size and confidence intervals, are currently being emphasized as a complement to null hypothesis statistical testing (NHST). However, introducing the new statistics to accompany NHST in the classroom has been a gradual process. This slow change may be the result of instructors not knowing how to begin teaching the new statistics in the classroom or a fear that changing their courses may be difficult. In this teacher-ready theory review, we begin by providing a brief overview of NHST and the problems associated with it; we then explain the new statistics and their advantages. We also offer two steps to help introduce the new statistics in the classroom that make calculating, interpreting, and discussing them easier for both instructors and students alike. We conclude with a discussion of resources that can provide more education regarding the news statistics and assist instructors with preparing their introductory statistics courses.

Keywords: the new statistics, statistical analysis, pedagogy, significance testing

In recent years, there has been a push for psychological science to rely less on the use of null-hypothesis statistical testing (NHST), which includes inferential statistics such as tests for analysis of variance (ANOVA) and *t* tests, and more on what are called the new statistics (Calin-Jageman, 2018; Cumming, 2014; Cumming & Calin-Jageman, 2016). The new statistics include, but are not limited to, measures of effect size, confidence intervals, likelihood ratios, and meta-analyses. The term the new statistics is a misnomer because these statistics have been around for decades, as documented by other scholars (e.g., Cohen, 1990; Goodman, 2008; Wilkinson & Task Force on Statistical Inference, 1999)—in fact, measures such as effect sizes were referenced in statistical texts as early as 1951 (Yates, 1951). But even though these statistics are not new, researchers still seem reluctant to use and report on these techniques. There are many reasons for this unwillingness. For example, with effect sizes, these rationales include (a) arguing that the information provided is redundant when reported with other inferential statistics (e.g., *p* values), (b) arguing that the reader should be able to calculate effect sizes on their own based on the statistics provided, and (c) arguing that journals do not require reporting of effect sizes and so they are unimportant (Washburn et al., 2018). These rationales provided by actual researchers are evidence that the new statistics are misunderstood in terms of what they add to research.

One indication of the renewed focus on the use of the new statistics can be seen with the recent ban of the use of NHST in the journal *Basic and Applied Social Psychology* (Trafimow & Marks, 2015), a ban that may portend a larger shift in the field toward revised statistical reporting standards. Indeed, it appears that psychological science is moving toward statistics beyond NHST in publications by encouraging researchers to use the new statistics in their reporting and providing free online tutorials

This article was published Online First April 1, 2019.

Christopher J. Normile, Emily K. Bloesch, Christopher C. Davoli, and Kyle C. Scherr, Department of Psychology, Central Michigan University.

Correspondence concerning this article should be addressed to Christopher J. Normile, Department of Psychology, Central Michigan University, Sloan Hall 101, Mount Pleasant, MI 48859. E-mail: normi1cj@cmich.edu

(Association for Psychological Science, 2014b; Cumming, 2014). Therefore, it is becoming important now more than ever that students learn how to compute and interpret these statistics in order to fully understand and evaluate scientific psychological research.

Although this content change may seem onerous to instructors currently teaching NHST, our goal in the current review is to alleviate those concerns by justifying the inclusion of the new statistics and demonstrating how they can easily be incorporated into introductory statistics classes. To accomplish this, we will provide (a) a description of the history of and problems associated with NHST, (b) a primer on the new statistics and their benefits, (c) practical steps for how to introduce the new statistics in the classroom, and (d) resources for education and implementation.

History of and Problems With NHST

NHST entered mainstream use when Fisher published *Statistical Methods for Research Workers* (Fisher, 1925). The manual was met with moderate excitement from researchers, as ANOVAs in particular gave insight into testing differences between groups (rather than their similarities with correlations) and could calculate estimated errors (Yates, 1951). In addition to explaining new statistical techniques, Fisher's manual stressed the importance of larger sample sizes in experiments, randomization of experimental and control groups, and replication.

However, since the introduction of NHST, researchers have argued that it is illogical and unintuitive (Cumming, 2014; Goodman, 2008; Wasserstein & Lazar, 2016). Moreover, it is believed that NHST fails to provide the answers regarding the information researchers really want to know, such as the probability that the alternative hypothesis is correct (Cohen, 1994). For example, a *p* value less than .05 tells us "assuming H_0 is true and the study is repeated many times, less than 5% of these results will be even more inconsistent with H_0 than the observed result" (Kline, 2004, p. 63). In other words, there is a less than 1 in 20 chance of finding a result more extreme than the one we found assuming the null hypothesis is true. Consequently, the *p* value tells us nothing about the magnitude of the effect of the independent variable or the likelihood that the alternative hypothesis is correct as the *p* value always assumes the null to be true (Cohen, 1994).

Other critics have argued that the problem with NHST is not the tests themselves but the inability of researchers to appropriately interpret the outcomes (Hagen, 1997). According to Fisher, an experiment that yields a small *p* value simply indicates that an experiment is worth replicating (Fisher, 1925). However, the meaning of the *p* value has been misconstrued over time. Indeed, the Task Force for Statistical Inference, a group created in 1999 by the American Psychological Association (APA) charged with uncovering the controversial issues regarding the application of statistics, had to stress to researchers that hypothesis testing was not just a dichotomy where they either accept or reject a decision. Rather, researchers should instead use their own judgment to (a) determine if the result from a computer-generated data analysis makes sense or not and (b) attempt to replicate their findings should analyses yield a small *p* value (Wilkinson & Task Force on Statistical Inference, 1999).

The misinterpreted dichotomous nature of the *p* value is a problem that students seem to struggle with even today. In a survey of 277 medical students in residency, only 58.8% were able to correctly answer a multiple-choice question that required the interpretation of the meaning of a *p* value (Windish, Huot, & Green, 2007). Respondents who answered incorrectly surmised that the *p* value indicated the probability that their results would be seen again if the study were repeated, while some concluded that the *p* value was the probability that the alternative hypothesis is correct. One possible reason for these misunderstandings may be that students erroneously believe that a statistically significant finding is clinically important (Goodman, 2008), even though the *p* value in and of itself tells researchers nothing about the size or importance of an effect (Wasserstein & Lazar, 2016). Of course, the confusion surrounding the interpretation of *p* values is not solely the fault of the students; rather, much of the blame can be placed on the unintuitive nature of the *p* value (Cumming, 2014; Greenland et al., 2016). Again, students and researchers seem to struggle with the notion that the *p* value always assumes the null hypothesis to be true and, therefore, cannot speak directly to the

probability that the null hypothesis or alternative hypotheses are true (Greenland et al., 2016). As a result, it would be beneficial to students for their instructors to teach the more intuitive new statistics.

The New Statistics and Their Benefits

The new statistics include a variety of techniques such as estimates of effect size, confidence intervals, and the use of meta-analyses (Cumming, 2014). One issue that many of these tests have in common is the lack of a dichotomous decision, meaning there are not predetermined cutoffs when evaluating the practical significance of effect sizes or confidence intervals.[1] As a result, researchers have the ability to ask themselves not only "*is this effect significant*," as they traditionally would, but also "*how strong is the effect of my treatment/independent variable?*" In contrast to testing for significance, the magnitude of the effect of the treatment/independent variable is more likely to provide practical information. In addition to measures of estimation, likelihood ratios can test competing alternative hypotheses and provide information about the probability that the alternative hypothesis or null hypothesis is correct (Dixon, 2003). Given that likelihood ratios are generally covered in more advanced statistics classes, they are beyond the scope of the current paper; however, Glover and Dixon (2004) provide an excellent starting point for those interested in learning more.

Effect sizes are a way to quantify the differences between two or more groups. In a clinical sense, effect sizes will tell you how well a treatment worked (rather than the dichotomous "*did the treatment work?*"). Like z scores, measures of effect size allow researchers to determine proportions between groups. For example, imagine a researcher conducts a study to compare the performance in a statistics class among students who took notes longhand (control group) to students who took notes using a laptop (experimental group). She finds the average of both groups on an exam and calculates the effect size using Cohen's d. If the researcher found the effect size to be -0.73, she would conclude that the average person in the laptop group scored 0.73 standard deviations below those in the longhand group. Consequently, assuming a standard deviation of 10, students in the laptop group would score, on average, 7.3 points less on the exam than students who took notes by longhand. This provides more information with regard to what note-taking methods professors should encourage, rather than simply saying there is a significant difference between note-taking methods. Examples such as this can be used in the classroom to highlight the importance of effect size in a practical sense.

Unfortunately, estimates of effect size are not perfect, as they are based on only one sample. However, confidence intervals can be used to assess variability and account for margins of error. Additionally, based on the confidence intervals obtained in one study, the need for replication can be determined—confidence intervals with larger ranges indicate more of a need for replication, smaller confidence intervals less so. It is important to note that the 95% confidence interval for effect sizes allows us to reason that, assuming we replicate our study a large number of times, the true population effect size would appear in 95% of the calculated confidence intervals (Greenland et al., 2016). Imagine that the 95% confidence interval for the effect size above (i.e., -0.73) is $[-1.12, -0.34]$. That means that in 95 of 100 replications of this study, the true effect size for the population is between -1.12 and -0.34. Perhaps more important, we can conclude that the effect is almost certainly negative (i.e., students who take notes with laptops do worse than students who take notes by hand) and that the effect could be big or small. It is at this point that professors could then stress the importance of replicating research. These future replications will then provide additional effect sizes and confidence intervals that will get us closer to estimating the true population effect size. Professors who already discuss replications can emphasize that p values can vary widely across replication studies even if all other variables and

[1] Some researchers use the heuristic that if a confidence interval includes a zero, there is no difference between groups. However, this is a further example of an NHST-derived interpretation. A zero in the interval simply implies that there is not sufficient evidence to conclude a difference between groups in that particular sample. While it is possible that there are no differences between the groups, it is also possible that the confidence interval lacks precision (i.e., is too wide) and a larger sample size is required to increase power and, thus, the precision of the confidence interval.

the effect size are held constant, an issue not shared with confidence intervals (Cumming, 2008; Lai, Fidler, & Cumming, 2012). The focus on replication, rather than a focus on a single *p* value (a poor assessment of the strength and practicality of effects on its own), will provide students with the tools needed to be successful in psychological research.

Steps for Introducing New Statistics

Step 1: Supplementing NHST Instruction With New Statistics

One way to broach the topic of the new statistics is to explain to students the issues surrounding NHST and *p* values, including the issues raised previously. This is important because students may question why some professors focus on NHST and others are trying to redirect their attention to the new statistics. Overall, the shift to prioritizing the new statistics over NHST does not need to be an onerous one. Professors can still teach many of the same topics they have always taught (e.g., *t* tests, correlations, ANOVAs). In addition, they could continue to use the same textbooks they always have or switch to easy-to-use texts that focus on the new statistics for introductory students (Cumming & Calin-Jageman, 2016). The biggest change will be the shift in focus from NHST to measures of estimation.

Instructors can still introduce NHST, but in a way that allows them to compare and contrast NHST with other statistics like confidence intervals in order to show the benefits of using the new statistics over NHST. For example, Coulson, Healey, Fidler, and Cumming (2010) conducted a study in which they had recently published researchers from a variety of disciplines interpret results from two similar studies presented in either NHST format or confidence interval format to determine the effectiveness of a treatment. They found that 44.1% of individuals who read the results presented in confidence interval format, which made no reference to NHST at all (i.e., *p* values, significance, etc.), discussed NHST during their interpretation of the findings (e.g., participants mentioning that the data were not significant if the confidence interval contained a zero). These same individuals were also more likely to incorrectly interpret the results compared to the other researchers who discussed the results in terms of interval extent and overlap. Therefore, the researchers who avoided thinking in terms of NHST were better at interpreting data than those who invoked such thinking. However, it is important to note that merely reporting confidence intervals when discussing results is not sufficient. Instead, it would behoove researchers to learn how to understand, interpret, and communicate confidence interval data and effect sizes without resorting back to terms found in NHST. For many, that learning will likely develop as a result of preparing to teach statistics in the undergraduate classroom.

Emphasizing the new statistics in introductory courses also has the benefit of allowing researchers and students alike to practice using critical thinking skills to determine whether an effect in their study is clinically or practically important. Currently, many researchers use and teach the Cohen benchmarks for effect sizes. That is to say, a score of 0.20 in Cohen's *d* is a small effect, 0.50 is a medium effect, and 0.80 is a large effect. However, it is important for researchers to use their own judgment when making a conclusion rather than relying solely on conventional benchmarks (Cumming, 2014; Thompson, 2002), a sentiment shared by Cohen (1988) himself. This will allow researchers and students some autonomy and once again allow them to think logically, rather than feeling the need to accept the interpretation of their data consistent with predetermined guidelines.

Step 2: Adopting New-Statistics-Friendly Software

It is well known that many students have anxiety and other negative feelings toward statistics courses (Garfield & Ahlgren, 1988; Wilson & Rosenthal, 1993). This anxiety can be exacerbated when students discover that they must learn how to use statistical software. Hence, some researchers have argued that instructors should not teach using statistical software in an introductory statistics class, as students tend to feel ambivalent toward the programs (Rosen, Feeney, & Petty, 1992). Others argue that students have more positive feelings toward statistics when they have a better understanding of the use of statistics in the real world (Pfannkuch & Wild, 2004; Yilmaz, 1996)—including, but not limited to, learning

how to use statistical software like researchers in the real world. In fact, more recent research has shown that students whose statistics courses include learning how to use statistical software had better grades than students whose courses used no such programs (Basturk, 2005). Therefore, to alleviate student anxiety and improve understanding, easy-to-use software that can easily give students what they need should be provided.

As the push for the new statistics increases, demand for software that can easily compute estimates of effect size and confidence intervals will also increase. One of the most dominant software programs on college campuses, the Statistical Package for the Social Sciences (SPSS), makes it difficult to find estimates needed for the new statistics as many of these measures, including Cohen's *d*, can only be calculated using syntax. Thus, if we want the new statistics to flourish, we need to introduce software that makes it easier for students to calculate these estimates. One such program is JASP, a free, open-source statistical software program that can be downloaded from the host website (JASP, 2018a). As an open-source program, students can have access to JASP from their personal computers without having to pay a large fee or do their work on a campus computer. One benefit of JASP is the fact that it can open files from Excel and SPSS. In this way, students can use the more intuitive Excel to create and label their variables and then import the data into JASP. The same can be done for students and professors who already have knowledge and skills using SPSS.

One of the most striking advantages of JASP is its ability to easily perform measures of estimation such as effect sizes and confidence intervals. For example, when conducting an ANOVA, users of JASP have the option of selecting from a variety of measures of effect size. This, in turn, allows researchers to have more control over the measure of effect size they believe works best for their research. In addition to effect size, researchers and students can add 95% confidence intervals to their tables and graphs by simply checking a box. Not only can they add confidence intervals for means, but they can also easily add confidence intervals for measures of effect size.

Another benefit of using JASP in class is the use of note-taking on output. These notes may include information such as how to write APA style results for students and, critically, how to present the new statistics. For example, after calculating an ANOVA, teachers and students can communicate their findings in APA style above or below the output table. These tables and notes can then easily be pasted into a word processor document for report writing.

Preparing Your Courses for the New Statistics

Online resources are available to help smooth the transition to focusing on the new statistics. These include six videos provided by the Association for Psychological Science that outline these concepts (Association for Psychological Science, 2014a). These videos can provide a refresher for instructors who have not encountered the new statistics recently. The first two videos focus on why using the new statistics is important, including information that introduces viewers to estimation and how the new statistics can improve research integrity. The last four videos provide information on how to properly conduct and interpret the new statistics, including effect sizes, confidence intervals, and meta-analyses.

For professors who wish to use a textbook that focuses on the new statistics, Cumming and Calin-Jageman (2016) released an introductory statistics book that highlights the importance of confidence intervals and effects sizes as well as open science and replicability. The textbook also covers NHST and allows for a dialogue between instructors and students to discuss the differences between the two approaches. In addition, the textbook contains in-chapter quizzes and exercises along with end-of-chapter exercises that use real data for student practice. The companion software, Exploratory Software for Confidence Intervals, can be used to assist students in visualizing estimation methods as well as create graphs and analyze data centered on confidence intervals.

For instructors who are unfamiliar with the statistical software JASP, a multitude of instructional resources exist. For example, the JASP website provides instructional videos on how to run a variety of tests, including *t* tests, ANOVAs, and Bayesian analyses (JASP, 2018b). Furthermore, JASP has a section of its website devoted to providing teaching resources to assist instructors with teaching their students how to use JASP (JASP, 2018c). These re-

sources include a data library and teaching materials such as PowerPoint slides for lectures and example assignments.

Conclusions

NHST is a statistical method that researchers and journals are beginning to deemphasize in favor of the new statistics (Cumming, 2014; Trafimow & Marks, 2015). Research has shown that researchers who avoid thinking in terms of NHST are better at interpreting data than those who invoke such thinking (Coulson et al., 2010). Therefore, it would be beneficial for teachers to highlight the new statistics in the classroom to best prepare students to understand and interpret research in the field. Moreover, researchers will benefit from teaching the new statistics as they too will need to know how to analyze and interpret these statistics in order to meet the evolving statistical reporting standards of many journals. These changes may seem laborious; however, the changes to already established introductory statistics courses do not have to be major. A straightforward way to begin the process is to integrate the new statistics concepts with traditional NHST content. For example, instructors can add information on confidence intervals and effect sizes to any of the inferential analyses they currently teach. In other words, instructors do not need to rework their entire class and can add the new statistics to supplement their current courses. This, in turn, will help students understand these concepts more readily. By incorporating these relatively straightforward additions to statistics courses, instructors will be able to provide students with the most up-to-date tools for understanding and conducting research as they continue their careers in psychology.

References

Association for Psychological Science. (2014a). *The new statistics: Estimation and research integrity*. Retrieved from https://www.psychologicalscience.org/members/new-statistics

Association for Psychological Science. (2014b). *Psychological Science sets new standards for research reporting*. Retrieved from https://www.psychologicalscience.org/news/releases/psychological-science-sets-new-standards-for-research-reporting.html

Basturk, R. (2005). The effectiveness of computer-assisted instruction in teaching introductory statistics. *Journal of Educational Technology & Society, 8,* 170–178.

Calin-Jageman, R. J. (2018). The new statistics for neuroscience majors: Thinking in effect sizes. *Journal of Undergraduate Neuroscience Education, 16,* E21–E25.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I've learned (so far). *American Psychologist, 45,* 1304–1312. http://dx.doi.org/10.1037/0003-066X.45.12.1304

Cohen, J. (1994). The earth is round ($p <. 05$). *American Psychologist, 49,* 997–1003. http://dx.doi.org/10.1037/0003-066X.49.12.997

Coulson, M., Healey, M., Fidler, F., & Cumming, G. (2010). Confidence intervals permit, but don't guarantee, better inference than statistical significance testing. *Frontiers in Quantitative Psychology and Measurement, 1,* 1–9. http://dx.doi.org/10.3389/fpsyg.2010.00026

Cumming, G. (2008). Replication and *p* intervals: *P* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science, 3,* 286–300. http://dx.doi.org/10.1111/j.1745-6924.2008.00079.x

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7–29. http://dx.doi.org/10.1177/0956797613504966

Cumming, G., & Calin-Jageman, R. J. (2016). *Introduction to the new statistics: Estimation, open science, and beyond*. New York, NY: Routledge. http://dx.doi.org/10.4324/9781315708607

Dixon, P. (2003). The p-value fallacy and how to avoid it. *Canadian Journal of Experimental Psychology, 57,* 189–202. http://dx.doi.org/10.1037/h0087425

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, UK: Oliver and Boyd.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19,* 44–63. http://dx.doi.org/10.2307/749110

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review, 11,* 791–806. http://dx.doi.org/10.3758/BF03196706

Goodman, S. (2008). A dirty dozen: Twelve p-value misconceptions. *Seminars in Hematology, 45,* 135–140. http://dx.doi.org/10.1053/j.seminhematol.2008.04.003

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations.

European Journal of Epidemiology, 31, 337–350. http://dx.doi.org/10.1007/s10654-016-0149-3

Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *American Psychologist, 52,* 15–24. http://dx.doi.org/10.1037/0003-066X.52.1.15

JASP. (2018a). *Download JASP*. Retrieved from https://jasp-stats.org/download/

JASP. (2018b). *How to use JASP*. Retrieved from https://jasp-stats.org/how-to-use-jasp/

JASP. (2018c). *Teaching with JASP*. Retrieved from https://jasp-stats.org/teaching-with-jasp/

Kline, R. B. (2004). What's wrong with statistical tests—And where we go from here. In *Beyond significance testing: Reforming data analysis methods in behavioral research* (pp. 61–91). Washington, DC: APA Books.

Lai, J., Fidler, F., & Cumming, G. (2012). Subjective *p* intervals: Researchers underestimate the variability of *p* values over replication. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 8,* 51–62. http://dx.doi.org/10.1027/1614-2241/a000037

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17–46). New York, NY: Kluwer Academic Publishers. http://dx.doi.org/10.1007/1-4020-2278-6_2

Rosen, E. F., Feeney, B., & Petty, L. C. (1992). An introductory statistics class and examination of SPSS/PC. *Behavior Research Methods, Instruments, & Computers, 26,* 242–244. http://dx.doi.org/10.3758/BF03204629

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect size. *Educational Researcher, 31,* 25–32. http://dx.doi.org/10.3102/0013189X031003025

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology, 37,* 1–2. http://dx.doi.org/10.1080/01973533.2015.1012991

Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., . . . Carsel, T. S. (2018). Why do some psychological researchers resist adopting proposed reforms to research practices? A description of researchers' rationales. *Advances in Methods and Practices in Psychological Science, 1,* 166–173. http://dx.doi.org/10.1177/2515245918757427

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70,* 129–133. http://dx.doi.org/10.1080/00031305.2016.1154108

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals. *American Psychologist, 54,* 594–604.

Wilson, W. C., & Rosenthal, B. (1993). Anxiety and performance in an MSW research and statistics course. *Journal of Teaching in Social Work, 6,* 75–85. http://dx.doi.org/10.1300/J067v06n02_07

Windish, D. M., Huot, S. J., & Green, M. L. (2007). Medicine residents' understanding of the biostatistics and results in the medical literature. *Journal of the American Medical Association, 298,* 1010–1022. http://dx.doi.org/10.1001/jama.298.9.1010

Yates, F. (1951). The influence of statistical methods for research workers on the development of the science of statistics. *Journal of the American Statistical Association, 46,* 19–34. http://dx.doi.org/10.2307/2280090

Yilmaz, M. R. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistics Education: An International Journal on the Teaching and Learning of Statistics, 4,* 1–9.

Received August 8, 2018
Revision received November 29, 2018
Accepted February 3, 2019 ■