The Pennsylvania Psychologist

Vol. 76, No. 1

JANUARY 2016 • UPDATE

Science and Psychology: Emerging Trends, Continuing Challenges

Samuel Knapp, EdD, ABPP; Director of Professional Affairs

Dr. Samuel Knapp

Never has the opportunity for meaningful psychological research been so great. Advances in information exchange, new sources of information (access to new subject pools and aggregate data sources called Big Data), and the new applications of statistical techniques have led to possibilities for the advancement of psychological science at a level not imagined even 20 years ago.

Most psychological research is still done with college students. However, new sources of participants include community samples from Mechanical Turk (a group of online community members who volunteer for various jobs including participating in psychological research). Also, researchers now have access to very large data sets (Big Data) they can use to find statistical trends that smaller populations or samples could not yield. For example, the National Institute of Mental Health used the extensive data that the Army had on its 500,000 soldiers to discern patterns that predicted successful suicide; that data was used to target more precise intervention programs (Sukel, 2015). Furthermore, researchers have access to data from public postings on social media. For example Eichstaedt et al. (2015) looked at the language found on Twitter messages indicating negative emotions (especially anger) and disrupted social relationships. This data predicted death from heart diseases in a community better than the CDC model using 10 demographic factors known to be related to heart disease. Finally, mobile electronic devices have allowed for ambulatory and contemporaneous assessment of an individual's thoughts, feelings, and even physiological states throughout the day.

These opportunities for expanded research populations and methodologies are occurring at a time when the validity of much published research is being questioned. Systematic flaws in the rewarding of scientists and the peer review process have led to the publication of much literature that is either inaccurate or inadequately qualified. For example, the Open Science Collaboration (2015) was able to replicate less than 50% of the significant findings reported in recent psychological literature. The issue is not unique to psychological science. In the field of

Continued on page 8

Culture and Research

Samuel Knapp, EdD, ABPP
Director of Professional Affairs

For most of its history, Westerners (Europeans or European Americans) have conducted most of the world's psychological research using Western subjects. Even among Westerners, English-speaking psychology has been especially dominant. Now that psychology has become more global, psychologists are increasingly questioning the extent to which findings with Western subjects represent universal principles that cut across all cultures.

Many findings from research done with Western populations were not replicated when done with persons from non-Western countries. As noted in the accompanying articles, some of this may occur because the original studies contain Type I errors (report a finding as significant when it is not). However, even well replicated studies with Western populations might not apply to non-Western cultures. In part this may be because the research participants are Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies; Henrich, Heine, & Norenzayan, 2010). Furthermore, Lalonde et al (2013) have cautioned that researchers

Continued on page 8

The Pennsylvania Psychologist
5925 Stevenson Avenue, Suite H, Harrisburg, PA 17112
www.papsy.org

Editor Tracie L. Pasold, PhD
PPA President Beatrice R. Salter, PhD
PPF President Pauline Wallin, PhD
Executive Director Krista Paternostro Bower, CAE

Your life now may be very different than it was ten years ago...

Has your financial planning changed to fit your current or future picture? A new or expanding business, mortgages, automobiles, a larger family… these can all contribute to a very different picture of your financial responsibilities today.

Group Term Life Insurance

Term Life Insurance can play an important role in your family's continued financial security should you die prematurely. Whether you need initial coverage or want to add to what you have, Trust Group Term Life Insurance[1] is affordable and has the features you will need to keep pace with changing family and financial responsibilities.

Call us at **1-877-637-9700** or visit **trustinsurance.com** for a no-obligation consultation.

Great Coverage at Affordable Premiums Including These Features:

- **Inflation Safeguard —** designed to prevent changes in the cost of living from eroding your death protection.[2]
- **Living Benefits —** allows early payment of death benefits if you become terminally ill.
- **Disability Waiver of Premium —** waives your premium payment if you become totally disabled.

THE TRUST
www.trustinsurance.com

[1] Available in amounts up to $1,000,000. Coverage is individually medically underwritten. Policies issued by Liberty Life Assurance Company of Boston, a member of the Liberty Mutual Group. Plans have limitations and exclusions, and rates are based upon attained age at issue and increase in 5-year age brackets.

[2] Inflation Safeguard offers additional insurance coverage and the premium will be added to your bill.

Bad Science?

Samuel Knapp, EdD, ABPP; Director of Professional Affairs

Steen (2011) found an increase in the detection of scientific fraud in recent years, although this may be an artifact of better techniques available to discover fraud. Psychological science is not immune, as two prominent researchers (one in social psychology and the other in non-human primate research) acknowledged reporting data that was either fabricated or collected without sufficient regard for its accuracy. Although I have no doubt that many findings have been deliberately fabricated, such fraud is probably less of a problem than the use of methodologies that mislead the readers (and perhaps the researchers themselves).

Many published studies lack sufficient sample size or have other methodological shortcomings that limit the confidence that should be placed in the reported results. The impact and pervasiveness of these flaws has been highlighted by recent efforts to replicate previous studies in the psychological sciences. A large portion of published studies have not been replicated.

Perhaps science is not as self-correcting as some once assumed (Iaonnidis, 2012). The self-correcting theory of science is that the insistence on statistical significance, critical reviews by blind peer reviewers, and findings by subsequent researchers would greatly limit the impact of any published finding that turned out to be inaccurate. However, do methodological errors occur far more frequently than previously believed and do false ideas, once published, perish as easily as some would believe? Does the culture of science sufficiently reward good science or significant results (even if obtained through less than optimal scientific conditions)? Scientists get grants, tenure, speaking engagements, and promotions if they publish findings with significant results; not if they use exact methodology.

Consider, for example, some common practices among drug researchers. If researchers looking at a drug find that the results approach, but do not reach, statistical significance, they can simply add more subjects until statistical significance is reached. Also, they can report only those dependent measures where significance was found. So if the study found that statistical significance was reached on only two of the four measures, the study can report only the two where significance was reached.

In addition, researchers can selectively recruit persons who have

Continued on page 4

Continued from page 3

participated in similar research studies in the past. Such participants often self-select themselves into these studies because they get free medications and they have a personal history of responding well to similar medications. On the surface it may appear that the researchers used a representative community sample; but they did not.

It is not uncommon for researchers to throw out statistical outliers or responses that deviate substantially from the normal responses. For example, one respondent may have been functionally illiterate and responded in near random fashion to the test questions. However, decisions as to what constitutes an outlier are not typically reported, thus allowing for the possibility that some of the outlier responses were, indeed, valid ones that researchers should have included in the data analysis.

In addition, the researchers may use numerous statistical analyses, including those not considered prior to conducting the experiment, to search for statistically significant results. Such "p-hacking" (Simonsohn, Nelson, & Simmons, 2014) procedures increase the likelihood that something significant would be found, but they also increase the risk that the significant finding was a random pattern without any theoretical or practical importance.

Finally, the problem is exacerbated by the "file-drawer effect" whereby a study that failed to find statistical significance was not even published, because the authors knew that journals are more interested in significant results and less interested in sound methodology.

So none of these decisions involved fraud in the sense that researchers fabricated data. But these practices, represent a series of research decisions which would allow the average reader (or perhaps the researchers themselves) to conclude that the drug had clinical effectiveness when, in fact, the clinical effectiveness was far less than the study implied or was even non-existent. Often researchers who make these decisions do not deliberately attempt to deceive the public; sometimes they believe so strongly in the hypotheses they are testing that they fail to appreciate the cumulative impact of these methodological decisions on the outcomes.

References

Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. ***Perspectives on Psychological Science, 7*****, 645–654.**

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. ***Journal of Experimental Psychology: General, 143*****, 534–547.**

Steen, R. G. (2011). Retractions in the scientific literature: Is the incidence of research fraud increasing? ***Journal of Medical Ethics, 37*****, 249–253.**

Restoring the Credibility of Psychological Research

Samuel Knapp, EdD, ABPP; Director of Professional Affairs

Several proposals have been made to correct some of the publication problems that currently exist in psychological research. These proposals include publishing studies that replicate previous findings; requiring the preregistration of research protocols; requiring data sharing; and giving sufficient publication space to well conducted experiments, even if the study does not find statistical significance. Many of these proposals have been around for a long time, although they are taking on new importance.

Replication of Previous Studies

Reproducibility is a sensitive issue. Senior researchers who spent years gathering information, doing literature reviews, conducting pilot studies, and finally publishing their research may resent having their methods and conclusions come under scrutiny. At its worst the replication process has an air of an Old West gunfight to it, wherein a young researcher rides into the town and seeks to bring down the old sheriff with a stellar reputation. But the replication project described below tried to minimize this potential competition. It urged researchers to work closely with the original authors in the methodology, design of materials, and other day-to-day elements of the design.

In 2011, the Center for Open Science assembled an international network of psychological scientists who attempted to replicate previously published studies in psychological journals ("Reproducibility Project"). Of the first 100 studies that were replicated, less than one half confirmed the positive results found in the original study. The authors noted that no one standard can be used to determine whether a replication was a success. When looking at probabilities (*p* values), 36% of the replicated studies showed a statistically significant effect. When looking at effect size, 47% of the replicated studies showed a statistically significant effect (Open Science Collaboration, 2015). Studies that were replicated successfully were more likely to be published in cognitive psychology as opposed to social psychology journals and the replicated studies were more likely to show more robust findings in the original study.

The implications of this project are that much of what is assumed to be true in psychological science might, in fact, not be true. Or, what was assumed to be true was only true for some populations or only under certain conditions. For those who understand the scientific process well, this should not be surprising. The progress of science necessarily involves efforts to falsify hypotheses.

> Reproducibility is a defining feature of science . . . Scientific claims should not gain credence because of the status or authority of their originator but by the replicability of their supporting evidence (Open Science Collaboration, p. 943).

It is possible that the findings from many of the original studies were inaccurate. But that conclusion should not be drawn from the results of one failed replication. First, replication is never completely perfect. No study can be replicated with complete fidelity; there will always be some differences (albeit sometimes minor) in the populations from which the experimental sample was drawn, and minor differences in actual experimenter behavior could

Continued on page 6

Continued from page 5

influence the results in some way. Also, the replication study may be a false negative (failure to find a result that actually exists). As summarized by the Open Science Collaboration:

> How many of the effects have we established are true? Zero. How many of the effects have were established are false? Zero. Is this a limitation of the project design? No. It is a reality of doing science. (p. 950)

Registering Study Designs

The Open Science Collaboration reported on a ground-breaking initiative, but it only represents the most prominent example of a movement toward greater transparency in science. For example, the journal *Experimental Psychology* has a registered report format in which authors of replication studies submit their proposal and design to the journal before the actual implementation of the study. "The paper will be published regardless of the results, given that the research is conducted and analyzed in accordance with the proposal and the conclusions follow from the results" (Stahl, 2015, p. 1). When studies are preregistered, researchers are more likely to report on hypotheses that were not substantiated or parts of the study that were inconsistent with the major hypotheses. Similar projects have been instituted by other journals. Medical researchers with government grants must preregister their methodology with clinical trials.gov, a registry of funded human subject research trials.

Data Sharing

In addition, many journal editors now require authors to share their data. Recent technological changes have eliminated many problems once associated with retaining and sharing data. One of the benefits of open data is that it may allow for the reanalysis of the original data set. Using data from medical trials, Ebrahim, et al. (2014) found that reanalyses sometimes led to changes in the direction of the findings and differences in the interpretation of the data with implications for patient care.

Publishing Non-Significant Results

Finally, some researchers are calling on journal editors to value methodologically sound articles that did not find significant results. Many journal publishers are interested in new and ground breaking studies and will often reject an article that is not original enough or which failed to find significant results, even if the study used good research methodology (Nosek & Lakens, 2014). Right now the publication system "incentivizes quantity more than quality, and novelty more than reliability" (Ioannidis et al., 2014, p. 166).

This priority for significant results may inadvertently degrade the quality of published research. For example, minor confounding variables can have a large influence on studies with small sample sizes, so there is a greater risk that significance will be found in studies with small sample sizes, even if the findings are false. This is not to imply that all studies with small samples are bad; only that they risk overestimating the effect size and need to be interpreted with caution (Ioannidis, Pereira, & Horowitz, 2013).

References

Ebrahim, S., Sohani, Z., Montoya, L., Agarwal, A., Thorlund, K., Mills, E. J., Ioannidis, J. P., et al. (2014). Reanalyses of randomized clinical trial data. *Journal of the American Medical Association, 312*, 1024–1032.

Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association, 294*, 218–228.

Ioannidis, J. P., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D. et al. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet, 383*, 166–175.

Ioannidis, J. P., Pereira, T. V., & Horwitz, R. I. (2013). Emergence of large treatment effects from small trials—Reply. *Journal of the American Medical Association, 309*, 768–769.

Nosek, B., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*, 137–141.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716.

Stahl, C. (2015). Experimental psychology: New standards. *Experimental Psychology, 62*, 1–2.

Statistics, Beer, and Feminism

Samuel Knapp, EdD, ABPP; Director of Professional Affairs

Statistics is a form of problem solving that involves the summary, organization, and analysis of data. Like many graduate students in psychology I went through my statistics courses thinking of statistical significance and null hypothesis testing as Holy Grails that could not be questioned. They were fixed verities that would last for eternity. Now this stable standard is being questioned. Cummings (2014) urged replacing null hypothesis testing with effect size estimation. Dienes (2011) urges that researchers consider using Bayesian techniques instead of the traditional probability estimations. Wetzels et al. (2011) looked at 855 analyses using *t*-tests and found that Bayesian and traditional tests often agreed in significance, but the statistics differed 70% of the time when the results were between .01 and .05 levels of significance. They concluded that "the Bayesian approach is comparatively prudent, preventing researchers from overestimating the evidence in favor of an effect" (p. 291). Perhaps what is needed is a more flexible approach whereby different methods could be used, depending on the data and the research questions. The new research methodology does not represent "mere tweaks to business as usual, but substantial change that will require effort" (Cumming, 2014, p. 7).

Now I understand that statistics evolves like other aspects of science to address societal needs. For example, William Gosset (1876-1937), an employee of Guinness beer, developed the "*t*-test." Guinness hired the best graduates from Oxford and Cambridge so that it could to apply biochemical and statistical techniques to its brewing. They would need to determine, for example, which strain of barley produced which flavor or consistency in its beer. Guinness did not allow its employees to publish trade secrets and the development of this statistical test could have been considered a trade secret. Nonetheless, the company turned a blind eye when Gosset published his article on the "*t*-test" using the pseudonym "Student" (1908).

Florence Nightingale (1820–1910) was known for creating the profession of nursing and in promoting sanitary conditions in hospitals. Her pioneering work in statistics is less widely known. Her progressive parents defied the standards of the time and allowed their daughter to pursue advanced studies in many fields including mathematics. During the Crimean War (1854–1856) between Russia and England far more soldiers died from disease than from combat. This was a well-known phenomenon that occurred throughout most campaigns in military history. But the hospital where Nightingale was superintendent of nurses saw its mortality rate fall from 43% to 2% in 18 months after it adopted her hygienic methods.

After the Crimean War, Nightingale was adored by the British public, but she was personally tormented by the many needless deaths she witnessed in British hospitals as a result of poor sanitary conditions. She urged the British army to adopt her uniform sanitary standards. Many physicians and politicians strongly opposed the institutionalization of hygienic methods. Nonetheless, a Royal Commission was established to study the impact of sanitary conditions on the health of soldiers. In addition to data gathered during the Crimean War, Nightingale unearthed data showing that British soldiers who lived in barracks during peacetime had higher mortality rates than civilians who lived in nearby cities. Through hard work, persistence, and the use of statistics (including innovative uses of pie charts and similar visual displays), she and her colleague William Farr prevailed. In recognition of her statistical contributions, Florence Nightingale was the first woman elected to the British Statistical Society in 1858, at the age of 38!

References

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7–29.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290.

Student. (1908). The probability error of a mean. *Biometrika, 6*, 1–25.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J., Iverson, G., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t*-tests. *Perspectives in Psychological Science, 6*, 291–298.

SCIENCE AND PSYCHOLOGY

Continued from page 1

psychiatric research the methodological problems associated with drug studies are so substantial that some researchers have questioned the efficacy of many commonly used psychotropic medications. Ioannides (2005) estimated that 50% of the published findings on medications contain inaccurate conclusions.

We are not always certain about how much of the published research is really a Type I error (false positives; information assumed to be true when, in fact, it is not). We do not always know how much of research is an artifact of data mining (the indiscriminate use of statistical techniques to find significant trends in the absence of any a *priori* research hypothesis). We are not always certain how much is valid only for Western cultures and has no or limited validity to other ethnic groups.

The following articles raise issues about trends in psychological science including the generalizability of findings to non-Western cultures, the use of questionable (but not fraudulent) research practices in drug research, current initiatives to address problems in research, and finally some comments on new statistical techniques.

References

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., et al. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science, 26*, 159–169.

Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association, 294*, 218–228.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac 4716.

Sukel, K. (2015 April). Big data help prevent Army suicides. *Monitor on Psychology*, 55–57.

CULTURE AND RESEARCH

Continued from page 1

should not consider all WEIRD participants as identical in terms of culture or ethnicity. Consequently, much of what we know to be "true" about psychological processes among Westerners may not be "true" for persons living in non-Western cultures; for ethnic or cultural subgroups within Western cultures; or they may be true in limited circumstances.

Consider, for example, the search for universally useful parenting patterns. In capitalistic societies, research shows that the authoritative patterns of parenting endorsed by Baumrind (1971) appear to produce the best adjusted children. However, does this represent a universally good way to raise children? Some evidence would suggest it does. For example Chinese or Chinese-Americans appear to thrive when their parents follow authoritative child rearing practices (Kim et al. 2013). Nonetheless, the Chinese-American children used in this study were from urbanized areas of Mainland China (such as Hong Kong) where capitalist values are widespread. Caution should be used in extrapolating the parenting findings to rural areas of China (or other non-capitalist cultures) where more traditional values are pervasive. As van de Vijver (2013) has noted "parental socialization strategies match the relationship patterns that prevail in society" (p. 763). The optimal parenting strategy may vary according to the society in which the child lives.

Similarly, some data suggests that the five-factor model of personality represents universal factors. However, this assumption may not be true in all cases. For example the agreeableness construct found in the five-factor model may not capture the nature of relationships found in collectivist societies (van de Vijver, 2013). Other examples could be given with many other concepts well established in Western psychology.

Of course this does not mean that we throw away everything learned from Western psychology. What it means is that we cannot assume that the concepts will necessarily be generalized beyond the research participant pool used in the original study.

References

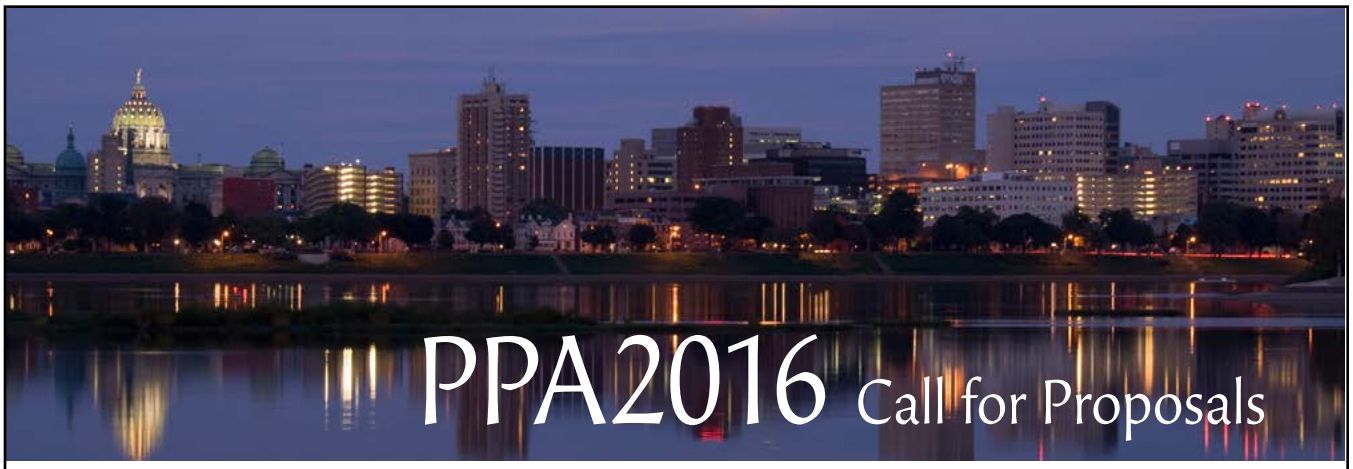
Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology, 4(1)*, 1-103.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Social Behavioral and Brain Sciences, 33,* 61–135.

Kim, S. Y., Wang, Y., Orozco-Lapray, D., Shen, Y., & Murtuza, M. (2013). Does "Tiger Parenting" exist? Parenting profiles of Chinese Americans and adolescent developmental outcomes. *Asian American Journal of Psychology, 4*, 7–18.

Lalonde, R. N., Cila, J., Lou, E., & Giguère, B. (2013). Delineating groups for cultural comparisons in a multicultural setting: Not all Westerners should be put in the same melting pot. *Canadian Journal of Behavioral Science, 45*, 296–304.

van de Vijver, F. J. R. (2013). Contribution of Internationalization to Psychology: Toward a global and inclusive discipline. *American Psychologist, 68,* 761–780.



PPA2016 Call for Proposals

Confronting Interpersonal Violence Throughout the Life Span

The 2016 PPA annual convention will address multiple areas of interpersonal violence, including the lasting impact of these experiences on the attainment of developmental stages, positive interpersonal relationships, and healthy coping skills. Areas that will be addressed include child abuse, domestic violence, intimate partner abuse, school violence including bullying, workplace violence, elder abuse, abuse of the disabled, and sexual abuse across the life span.

The goal is to establish strategies to help our communities move from cultures of detachment and violence to cultures of cooperation, peacefulness, and nonviolence. Topics will seek to identify ways to highlight psychology's value in homes, schools, workplaces, and our communities at large. We, as psychologists, can become change agents by working with other professionals and agencies to identify causes that fuel violence and develop concrete tools that address and remedy these issues. Nonviolence can be promoted through advocacy and the dissemination of educational tools that reinforce an evolving view of cultural differences, self-awareness, violence, and trauma.

PPA2016

Overcoming Interpersonal Violence Throughout the Life Span

All topics submitted by **January 8, 2016**, will be reviewed by the Convention Committee, but special consideration will be given to topics that relate to this year's theme. Submit your proposal online at http://www.papsy.org/?page=Convention.

Questions about this process?
Please contact Judy Smith at judy@papsy.org or (717) 510-6343

2016 PPA Continuing Education

PPA is looking to continue its long-standing tradition of offering high-quality CE programs to psychologists. In 2016 we are looking to expand these options – we hope you'll join us for one of these programs!

Calendar

Spring Continuing Education and Ethics Conference
April 8
Pittsburgh Area, PA

PPA on the Road
TBA

PPA2016 – PPA Annual Convention
June 15–18
Harrisburg, PA

Fall Continuing Education and Ethics Conference
TBA

PPA on the Road
TBA

Contact: judy@papsy.org

Webinars and Home Studies
Available year-round!

Podcasts
Podcasts for CE credit by Dr. John Gavazzi are available on www.papsy.org.

For CE programs sponsored by one of the Regional Psychological Associations in Pennsylvania, visit www.papsy.org.

Registration materials and further conference information are available at www.papsy.org.

Home Study CE Courses

Pennsylvania Child Abuse Recognition and Reporting: 2015 (Act 31 Approved) NEW!
2 CE Credits

Medicare's 2015 Physician Quality Reporting System (PQRS) NEW!
1 CE Credit

Excess Weight and Weight Loss
3 CE Credits

*Ethical Practice Is Multicultural Practice**
3 CE Credits

*Introduction to Ethical Decision Making**
3 CE Credits

Staying Focused in the Age of Distraction: How Mindfulness, Prayer, and Meditation Can Help You Pay Attention to What Really Matters
5 CE Credits

*Competence, Advertising, Informed Consent, and Other Professional Issues**
3 CE Credits

*Ethics and Professional Growth**
3 CE Credits

*Foundations of Ethical Practice**
6 CE Credits

*Ethics and Boundaries**
3 CE Credits

Readings in Multiculturalism
4 CE Credits

*Pennsylvania's Psychology Licensing Law, Regulations, and Ethics**
6 CE Credits

***This program qualifies for 3 contact hours for the ethics requirement as mandated by the Pennsylvania State Board of Psychology.**

For all Home Study CE Courses above contact: Judy Smith, (717) 510-6343, judy@papsy.org.