

## Do I know as much as I think I do? The Dunning-Kruger effect, overclaiming, and the illusion of knowledge

Nejc Plohl<sup>1\*</sup> and Bojan Musil<sup>2</sup>

<sup>1</sup>Ruše, Slovenia

<sup>2</sup>Department of Psychology, Faculty of Arts, University of Maribor, Slovenia

**Abstract:** Realistic perception of our own knowledge is important in various areas of everyday life, yet previous studies reveal that our self-perception is full of shortcomings. The present study focused on general overestimation of knowledge and differences between experts and the less-skilled (The Dunning-Kruger effect), self-perceived knowledge of non-existing concepts (overclaiming), and the illusion of knowledge. These phenomena were tested with an instrument which measured the actual knowledge of different domains (grammar, literature, and nanotechnology), as well as self-assessed knowledge. Results showed that, on average, participants overestimated their absolute performance, but not their performance relative to others. Furthermore, the bottom quartile overestimated their absolute and their relative performance most, while the top quartile perceived their absolute performance most accurately and substantially underestimated their relative performance. Results related to overclaiming showed that 56% of respondents claimed knowledge of at least one non-existent book and that the extent of overclaiming was substantially correlated with self-perceived expertise. Lastly, results showed that an increased quantity of information about nanotechnology led to a false certainty in answering questions from this area.

**Keywords:** overestimating knowledge, Dunning-Kruger effect, overclaiming, illusion of knowledge

## Ali znam toliko, kot mislim, da znam? Dunning-Krugerjev učinek, poznavanje neobstojećih konceptov in iluzija znanja

Nejc Plohl<sup>1\*</sup> in Bojan Musil<sup>2</sup>

<sup>1</sup>Ruše

<sup>2</sup>Oddelek za psihologijo, Filozofska fakulteta, Univerza v Mariboru

**Povzetek:** Realistično zaznavanje lastnega znanja je pomembno na različnih področjih vsakdanjega življenja, a pretekle raziskave razkrivajo, da je naša samozaznava v resnici polna pomanjkljivosti. Pričujoča študija se je osredotočila na splošno precejevanje znanja in razlike med eksperti ter slabšimi reševalci (Dunning-Krugerjev učinek), na poznavanje neobstojećih konceptov in na iluzijo znanja. Omenjene pojave smo preverjali z instrumentom, ki je vseboval elemente objektivnega ocenjevanja znanja in samoocene znanja na različnih področjih – slovnica, književnost in nanotehnologija. Rezultati so pokazali, da so udeleženci v povprečju precejevali svoj absolutni dosežek, ne pa tudi svojega položaja v vzorcu. Nadalje so tisti iz spodnjega kvartila najbolj precejevali svoj absolutni in relativni dosežek, medtem ko so tisti iz zgornjega kvartila svoj absolutni dosežek ocenjevali najtočneje, ob tem pa znatno podcenjevali svoj relativni dosežek. Rezultati, vezani na poznavanje neobstojećih konceptov, so pokazali, da je 56 % udeležencev zatrnilo poznavanje vsaj enega neobstojećega literarnega dela, pri tem pa se je z ravno poznavanja neobstojećih konceptov zmerno povezovala samozaznana kompetentnost. Nazadnje so rezultati pokazali tudi to, da je povečana kvantiteta informacij o nanotehnologiji vodila do občutka lažne gotovosti pri odgovarjanju na vprašanja iz tega področja.

**Ključne besede:** precejevanje znanja, Dunning-Krugerjev učinek, poznavanje neobstojećih konceptov, iluzija znanja

\* Naslov/Address: Nejc Plohl, Ruše, Kurirska pot 43, 2342 Ruše, e-mail: nejcplohl@gmail.com

Accurate perception of our own skills and knowledge is of utmost importance. A driver who is aware that he is not experienced enough to drive in certain situations will try to avoid them, or at least be more careful while driving in these circumstances; a medical doctor who knows that he is not competent to treat a particular disease, will refer a patient to a colleague who does possess this specific knowledge; and an individual who finds himself amid the debate about an unfamiliar topic, will make the wise decision to stay silent. Such reactions, all the result of accurate self-perception, would lead to greater road safety, better quality of treatment, and improved debates. They are, however, uncommon according to past research. Studies (e.g. McKenna, Stanier, & Lewis, 1991) showed that people generally believe that their driving is above average; this applies to overall competence, as well as to individual manoeuvres such as overtaking. Similarly, studies conducted on general practitioners showed weak correlations between self-assessed medical knowledge and objective knowledge (Tracey, Arroll, Richmond, & Barham, 1997). At last, various examples demonstrated that people do not always stay silent when they do not know something. In one of the episodes of the so-called “Lie Witness News” (part of the “Jimmy Kimmel Live!” TV show), a random passer-by was asked about a fictional band called “Tonya and the Hardings”; not only did the passer-by claim to know the band, she gave an elaborate response, saying that it is an all-female band that pushes the boundaries of the music industry (Dunning, 2014). These, rather specific examples, lead to a far more general conclusion: while accurate perception is indeed important and could be beneficial in many situations, we all have trouble assessing our own knowledge.

In the present study, an otherwise broad topic – self-assessment of knowledge and skills – is reduced to only a few interesting phenomena that highlight important deficiencies in self-perception. Hence, we focus on three main research questions. First, do people generally overestimate their knowledge, and is this overestimation typical of experts as well as the less-skilled? Second, how often do people claim to know something that they do not actually know, and what drives this type of behaviour? Third, how does increased quantity (but not quality) of information affect self-perception of knowledge on a specific topic? While these phenomena, in general, already have some empirical support, past literature still contains substantial gaps that need to be addressed to additionally support the existence of these lacunae and to truly understand mechanisms that underlie them. As such, our study introduces changes to the well-established procedures of measuring these phenomena (potentially increasing the validity of findings) and tests underlying mechanisms that are still in need of additional empirical examination (potentially adding support to explanations that are currently based on limited empirical findings). As opposed to most other studies which focused on only one phenomenon (e.g. only overclaiming), this study assesses three different phenomena on the same participants, thus allowing for a comparison between them. Furthermore, in contrast to previous studies which were almost exclusively conducted in English-speaking countries, we tested these phenomena on a sample of Slovene students.

## The Dunning-Kruger effect

In many domains of life, success depends on skills that allow us to follow “correct” strategies and rules (i.e. those that will help us achieve the desired results). These correct strategies are often domain-specific instead of general: effective management of a company, forming a solid logical argument, and planning a rigorous psychological study all require different competences (Kruger & Dunning, 1999). Since people differ in competences needed in particular domains, their outcomes in these domains differ as well (Dunning, Meyerowitz, & Holzberg, 1989). Of particular importance to the present study, Kruger and Dunning (1999) reported that when people are incompetent in the strategies they adopt to achieve success, they suffer a dual burden. First, they reach wrong conclusions and make unfortunate choices. Second, their incompetence robs them of the ability to realize it. Hence, they are left with the mistaken impression that they are doing just fine.

What mechanism lies behind the false impression of “doing fine”? Kruger and Dunning (1999) argued that the abilities which diminish competence in a particular domain are normally the same abilities that would be needed for accurate self-assessment in this domain. In cases where people do not recognize that they are wrong, it is hence reasonable to expect inflated judgments about their own performance. Trying to put this claim into a broader theoretical framework leads us to metacognition, a concept that encapsulates awareness of how well we are doing and how likely it is that our judgments are indeed accurate (Everson & Tobias, 1998).

Although the first relevant findings date more than 30 years back (e.g. Chi, Feltovich, & Glaser, 1981; Kunkel, 1983), this line of research largely emerged after a study by Kruger and Dunning (1999), who studied self-assessment in various areas: humour, logical reasoning, and, of particular relevance to our study, grammar. To this purpose, they distributed an instrument that included an objective test and a range of self-assessment questions. The results showed that respondents generally overestimated their ability relative to others as well as the number of correctly completed tasks. Additionally, Kruger and Dunning (1999) performed analyses by dividing participants into quartiles based on their performance. The first quartile, the quartile of participants who performed worst, placed their performance relative to others in the 61<sup>st</sup> percentile, even though their actual performance belonged in the 10<sup>th</sup> percentile. These participants also overestimated their absolute performance on the test. Participants from the second and third quartiles overestimated their performance considerably less than the bottom quartile, while those in the top quartile, the quartile of participants who performed best, even underestimated their knowledge; their actual achievement belonged in the 89<sup>th</sup> percentile, but they placed it in the 70<sup>th</sup> percentile. The top quartile, interestingly, did not underestimate their absolute performance on the test; instead, they assessed it rather accurately. A similar pattern emerged in other parts of the study; the only important difference appeared in the study of logical reasoning, where the top quartile underestimated both their relative performance as well as their absolute performance.

The study by Kruger and Dunning (1999) encouraged other scholars to study the effect, today widely recognized as the Dunning-Kruger effect. It has since been replicated in the field of grammar (e.g. Pavel, Robertson, & Harrison, 2012) and found in other domains, such as chemistry (Bell & Volkman, 2011; Pazicni & Bauer, 2014), information literacy (Gross & Latham, 2012), and emotional intelligence (Sheldon, Dunning, & Ames, 2014). Based on previous paragraphs, we predict the following:

- H1a:* Participants will, on average, overestimate their absolute performance.  
*H1b:* Participants will, on average, overestimate their relative performance.  
*H2a:* The bottom quartile will overestimate their absolute performance the most.  
*H2b:* The bottom quartile will overestimate their relative performance the most.  
*H3a:* The top quartile will estimate their absolute performance most accurately.  
*H3b:* The top quartile will underestimate their relative performance the most.

Despite many studies which demonstrated the Dunning-Kruger effect in various fields, there are, however, still some questions that need to be addressed. First, does the Dunning-Kruger effect remain when knowledge is examined thoroughly with different types of tasks? In fact, many previous studies have largely relied on relatively short tests with 10–20 similar multiple-choice questions (e.g. Kruger & Dunning, 1999; Pavel et al., 2012). Second, does the Dunning-Kruger effect remain when the task at hand is highly difficult? To our knowledge, not many studies tested the Dunning-Kruger effect with such exams. Additionally, studies that did, showed opposing results. In a study by Kruger and Dunning (1999), participants averagely attained 49.1–66.4% of correct answers, indicating a relatively high difficulty of the exam. Nevertheless, these authors report a vast overestimation of relative performance in the bottom quartile. On the other hand, Burson, Larrick, and Klayman (2006) reported poor performers being quite accurate when assessing their performance on a highly difficult task. In the present study, we thus aim to test our hypotheses on a thorough and high-difficulty grammar test.

## Overclaiming

The previous section leads us to the conclusion that our self-perception is often inaccurate and that there are key differences between experts and the less-skilled (Kruger & Dunning, 1999). In the following paragraphs, we will go one step further: in certain situations, people claim to know more than is possible, or, to put it differently, that they know concepts which do not exist. This phenomenon is called overclaiming (Atir, Rosenzweig, & Dunning, 2015).

The earliest observations of this phenomenon date back to at least the 1980s. In a study by Bishop, Oldendick, Tuchfarber, and Bennet (1980), almost one-third of respondents

expressed their opinion on the “1975 Public Affairs Act” – a completely fictitious law<sup>1</sup>. Researchers working on a recent series of polls “Public Policy Polling” (2015) stumbled across a similar finding; almost one-third of respondents supported the bombing of Agrabah – a completely fictitious country from the Disney animated movie *Aladdin*.

In addition, overclaiming has also been demonstrated in more rigorous laboratory studies, which are often designed so that the participants assess their familiarity with real and fabricated concepts from a particular domain (Atir et al., 2015; Paulhus, Harms, Bruce, & Lysy, 2003). In one of the most recent and detailed studies of this phenomenon, Atir et al. (2015) gave participants a list of 15 seemingly financial concepts and asked them to rate their knowledge of each concept with an appropriate value, ranging from 1 to 7. Twelve items represented real financial concepts (e.g. inflation), while the remaining three items were completely fictitious (e.g. annualized credit). In study 1a, 93% of the participants claimed to be at least somewhat familiar with at least one fictitious concept. The percentage of people who overclaimed was similarly high in study 1b (91%).

Even though the literature reporting a tendency to overclaim has accumulated in the last few years, very few studies have aimed to discover the mechanisms behind the phenomenon. Hence, the issue of when people are more likely to express this tendency is still largely unaddressed. One of the rare studies which tried to reveal the underlying mechanisms focused on the role of self-perceived expertise in a particular area (Atir et al., 2015). The presumptions of this study can be illustrated with an example: if John believes that his knowledge of biology is excellent, while Nathan, on the other hand, believes that his knowledge of biology is poor, John is more likely to say that he is familiar with fictitious biological concepts. Similarly, if John considers himself to be an expert in biology and less skilled in the field of philosophy, he is more likely to overclaim in the field of biology than in philosophy. While these claims are rather novel in relation to overclaiming, older studies have already shown some indirect support for the important role of self-perceived expertise (e.g. Bradley, 1981; Ehrlinger & Dunning, 2003). Additionally, studies 1a and 1b by Atir et al. (2015) indeed showed that self-perceived expertise positively predicted overclaiming; the more participants perceived themselves as competent in the field of personal finances, the more they claimed to know the fictitious concepts seemingly coming from this domain. At the same time, self-perceived expertise also correlated positively with actual knowledge. Moreover, the second study (1b but not 1a) revealed the order effect: overclaiming was more pronounced when participants assessed their self-perceived expertise before responding to the main overclaiming task (compared to

<sup>1</sup> Interestingly enough, similar ideas (i.e. “knowing” concepts that do not exist) can be found in Yugoslavian scientific literature that predates Bishop’s (1980) study; Golčić (1972) conducted a study on “semantic snobbism” and found that respondents recognized nonsensical but familiar-sounding words as legitimate foreign words.

the opposite order). However, self-perceived expertise was a statistically significant predictor of overclaiming in both situations.

Overclaiming has also been demonstrated by a few other studies. Swami, Papanicolaou, and Furnham (2011) examined overclaiming in the field of mental health, while Paulhus and Harms (2004) tested and further supported the phenomenon in regard to general knowledge. Based on these studies, we predict the following:

*H4:* The majority of participants will claim to be familiar with at least one fictitious concept.

*H5a:* There is a positive relation between self-perceived expertise and the number of familiar real concepts.

*H5b:* There is a positive relation between self-perceived expertise and the number of familiar fictitious concepts.

*H6:* Overclaiming will be higher when participants assess their competence before responding.

As indicated by the presented literature, studies on the relation between self-perceived expertise and overclaiming are relatively scarce. Additionally, to the best of our knowledge, only Atir et al. (2015) tested the order effects on overclaiming. Hence, we investigated these two effects, which could further illuminate our understanding of the phenomenon. Furthermore, all studies presented above tested overclaiming with a help of questionnaires which employed Likert scales. In contrast, we propose that Likert scales might not be particularly valid in this case; our perceived knowledge of fictitious and real concepts is unlikely to be that specific and differentiated (what is the difference between a 4 and a 5 when estimating how familiar you are with a fictitious book?). Hence, in our study, we employed a dichotomous response format in the main overclaiming task and investigated whether this would also result in overclaiming.

## Illusion of knowledge

The previous two phenomena emphasize that people often overestimate their knowledge and claim to know concepts that they do not know. The final phenomenon that we consider in this paper is strongly related: lack of knowledge does not necessarily lead to disorientation and confusion, but to a feeling of certainty about objectively inadequate knowledge (Dunning, 2014). False certainty can be very problematic. In the case of young drivers, Gregersen (1996) noted that overestimation of driving skills leads to a greater likelihood of involvement in a traffic accident. The logical solution to this false certainty is, at least intuitively, education, i.e. a process that can educate drivers about their limits and about difficult situations (Gregersen, 1996). However, when we step away from intuition and seek empirical support for the assumption that education necessarily leads to better skills and more accurate self-assessment, we stumble upon an interesting opposing fact: education often enhances the illusion of knowledge (Dunning, 2014).

Additional driver education courses have rarely been empirically proven to be fully effective. On the contrary, many studies showed that there are no positive effects on road safety (Gregersen, 1994). Among authors who associate the lack

of positive effects with training as such, there is considerable support for the explanation that trainees overestimate the effect of the training program. In other words, participants believe that their driving must be better since they have acquired a lot of new information (Gregersen, 1996). Such a reaction is by no means restricted to driving; Schwarz (2004) reported that people generally believe there is a positive linear relationship between more information and better decisions.

Hall, Arris, and Todorov (2007) empirically tested and supported the hypothesis that gaining more information often reduces the actual accuracy of predictions (predictions of uncertain outcomes) and simultaneously raises belief in the accuracy of these predictions. Several other studies have also shown that increasing the amount of information often increases certainty in judgments, even though the actual accuracy of these judgments does not change (e.g. Gill, Swann, & Silvera, 1998; Heath & Tversky, 1991). Based on these studies, we predict the following:

*H7:* Participants who receive more information about the tested knowledge domain will be more certain about the correctness of their answers.

In the present study, we aimed to replicate this phenomenon by including a topic that people are not very familiar with (nanotechnology) and by manipulating only the amount of irrelevant text. Additionally, we controlled for the initial familiarity with the topic. Furthermore, as opposed to the majority of previous studies which were conducted in the United States (e.g. Gill et al., 1998; Hall et al., 2007; Heath & Tversky, 1991), we tested this effect on a sample of Slovenian students.

## Aim of the study

In sum, the core aim of the present study was to investigate self-assessment of knowledge by testing several phenomena related to overconfidence (i.e. the Dunning-Kruger effect, overclaiming, and the illusion of knowledge) and mechanisms that underlie them. More specifically, we were interested in finding out whether these phenomena would emerge despite the modifications to the well-established ways of measuring them, and despite a non-traditional (non-English-speaking) sample. The present study also explored whether overestimation is a result of a general and stable trait or, in contrast, a rather task- and domain-specific phenomenon.

## Method

### Participants

The sample consisted of 91 participants, including 83 women and 8 men, who had an average age of 20.35 years ( $SD = 1.31$ ). All participants were undergraduate students of psychology or sociology. Since some participants failed to fill out certain parts of the instrument, they had to be excluded from the analyses that addressed those parts of the instrument; four participants were excluded in the first part, four participants in the second part, and three participants in the third part of our study.

## Instruments

Our instrument has two versions, both of which are in Slovene and composed of four parts; some of these parts include manipulations and therefore differ between the two versions. However, both versions start with a universal Part A, designed to obtain basic demographic data: gender, age, field and year of study.

Part B is designed to test the Dunning-Kruger effect. It consists of a grammar test and two self-assessment questions. The grammar test contains 27 tasks which are of different types: questions with two alternatives, multiple choice questions with four alternatives, tasks that require insertion and short answers, and tasks which require participants to read a sentence and determine whether it contains errors, and, if they deem it necessary, repair the sentence. Moreover, these tasks test a wide range of grammar knowledge, such as the correct use of commas and capital letters, declensions of nouns, finding a suitable synonym, etc. All tasks were – some directly, some in a slightly modified form – adopted from previous Slovene grammar Matura tests (i.e. high-school graduation tests). Participants were warned about the application of correction for guessing. The grammar test was followed by two further questions. Participants were asked to assess their achievement: first their absolute achievement (predicted percentage achieved on the test) and then their relative achievement by marking their score relative to others on a line (predicted percentile). Part B was same in both versions of our instrument.

Part C is designed to test perceived knowledge in the field of literature, particularly familiarity with the bibliography of the Slovene author Ivan Cankar. Participants were informed that there were no right or wrong answers and that we were only interested in their familiarity with certain works written by this author. The overclaiming task contains 12 items: 8 real (e.g. “Na Klancu”) and 4 fictitious works (e.g. “Naša zemlja”). Participants respond with either “Yes” (if they are familiar with this work) or “No” (if they are not familiar with this work). The order of real and fictitious items is random and equal for all participants. Besides the central overclaiming task, Part C also contains a question about self-perceived competence (“How familiar are you with the bibliography of Ivan Cankar?”) with a 7-point response format. Half the participants responded to this question before tackling the overclaiming task, while the other half responded to this question after completing the main part.

Part D is seemingly designed to test knowledge of nanotechnology, but in reality tests the illusion of knowledge caused by increased quantity (but not quality) of information. To test this phenomenon, we decided to use a topic which most people are unfamiliar with (or less familiar); by doing so, we were able to ensure that judgments about certainty would be susceptible to the information provided by us. At the beginning, participants had to answer a control question with a 4-point response format: “How much have you heard about nanotechnology until today?”. This question was followed by a passage of text, which was short and contained very little information in the control condition (one general paragraph on nanotechnology), while the text in experimental condition contained

more information (three paragraphs on nanotechnology: one general paragraph and two paragraphs about the benefits and risks of nanotechnology). The text about nanotechnology was adapted from a study by Kahan, Braman, Slovic, Gastil, and Cohen (2009) and did not contain any information that could be helpful in the short quiz that followed. As we have already indicated, the passage of text was followed by four questions on nanotechnology, e.g. “Who coined the term nanotechnology?”. All four questions were multiple choice questions with three alternatives. Additionally, each question contained a special supplementary question about participants’ certainty about the correctness of their answer (from 0 to 100%). The main purpose of these questions was to compare average certainty between the two experimental groups while making sure that the actual accuracy in both groups was always 0%; none of the alternatives were, in fact, correct.

## Procedure

The data was obtained collectively. Participants were randomly allocated to one of the two conditions; half of the participants completed the version 1 of our instrument (Part C: self-assessment before the overclaiming task, Part D: more information) and the other half completed the version 2 of our instrument (Part C: self-assessment after the overclaiming task, Part D: less information). In the recruitment phase, participants were guaranteed anonymity and reminded that their participation is completely voluntary. Completing the instrument took about 25 minutes. After the study, participants were briefly informed about the purpose of our study and encouraged to ask any questions. Statistical analyses were performed using Microsoft Excel 2016 and IBM SPSS Statistics 23.

## Analysis

Part B, which is a test of knowledge, was examined and scored according to the prepared criteria (in scoring, incorrect answers to alternative and multiple-choice questions were given negative points). The predicted score for each respondent, originally assessed in percentages (easier for participants), was transformed into points. Based on raw scores (absolute performances), a position in the sample (actual percentile) was calculated, and respondents were allocated into four quartiles. The predicted percentile was also entered into our database. Before analysing the data collected from Part C of our instrument, individual responses to each of the 12 items were entered into our database; “Yes” answers (indicating familiarity with a concept) were assigned one point. Based on individual responses, the number of familiar real and fictitious concepts was calculated. Preparing the responses from Part D for analysis required an additional calculation of average certainty in selected answers.

Using IBM SPSS Statistics 23 we then analysed the basic properties of all variables and checked the normality of variables’ distributions. To test our hypotheses, we used a variety of statistical tests. The first two hypotheses (*H1* and *H2*) were tested with the Wilcoxon Signed-Rank test. The next four hypotheses were tested using ANOVA and its post hoc

tests. The next few hypotheses were tested with correlations, specifically with Spearman's rho coefficient. The last two hypotheses (*H6* and *H7*) were tested with procedures that allow a comparison of two independent samples: *H6* with the Mann-Whitney *U* test and *H7* with the independent samples *t*-test (and the addition of ANCOVA). All statistical tests are accompanied by effect sizes (Cohen's *d* and  $\eta_p^2$ ).

## Results

### The Dunning-Kruger effect

First, we analysed whether individuals generally overestimate their absolute knowledge. The results showed that the predicted absolute score ( $M = 18.20$ ;  $SD = 5.65$ ) was higher than the actual absolute score ( $M = 13.11$ ;  $SD = 5.64$ ); the difference between the two was more than 5 points. These results also indicate that the test was, indeed, of high difficulty; on average, participants attained 48.6% of correct answers. In addition to descriptive analyses, we also performed the Wilcoxon Signed-Rank test, which showed that the average of positive ranks (expected > actual absolute score) was significantly higher than the average of negative ranks (expected < actual absolute score);  $Z = -6.39$ ,  $p < .001$ ,  $d = 0.90$ . Similar analyses were conducted for relative performance: do people generally overestimate their position in the sample? The results showed that the actual percentile ( $M = 50.57$ ;  $SD = 29.35$ ) turned out to be slightly higher than the predicted percentile ( $M = 48.92$ ;  $SD = 16.76$ ), but the two did not differ significantly;  $Z = -0.38$ ,  $p = .71$ ,  $d = 0.07$ .

In further analyses, participants were divided into four quartiles, based on their absolute score on the grammar test. Table 1 shows the average absolute scores, predicted absolute scores and differences between expected and absolute scores for each quartile. The average absolute score increased gradually from the first to the fourth quartile, with standard deviations being higher in extreme quartiles. Average predicted

absolute scores showed a similar pattern; they increased gradually from the bottom to the top quartile. Standard deviations, however, showed the opposite pattern (compared to actual absolute scores): variability was highest in the middle two quartiles. Actual and predicted absolute scores were necessary for the calculation of the difference between predicted and actual absolute performance. Results from Table 1 show that the difference between predicted and actual absolute performance was highest in the bottom quartile, the quartile containing the least skilled participants. The bottom quartile was followed by the second and third quartile respectively, while participants from the top quartile, the quartile of "experts", perceived their knowledge most accurately. The standard deviations were very similar in all quartiles; variability was only slightly lower in the extreme quartiles.

A one-way ANOVA showed that the difference between the predicted and the actual absolute score differs statistically significantly between groups,  $F(3, 83) = 5.71$ ,  $p = .001$ ,  $\eta_p^2 = .17$ . This result was first explored by comparing the bottom quartile with the remaining quartiles. As it turned out, the average difference between the expected and the actual absolute achievement was significantly higher in the first quartile compared to the fourth quartile ( $p < .001$ ), but not compared to the second ( $p = .66$ ) and third ( $p = .06$ ) quartiles. Additionally, Cohen's *d* occupied a low value when comparing the first and the second quartile ( $d = 0.14$ ); a medium value for the comparison between the first and the third quartile ( $d = 0.59$ ); and a high value for the comparison between the first and the fourth quartile ( $d = 1.18$ ). We performed identical analysis for the top quartile as well. The results showed that the difference between the expected and the actual absolute achievement was significantly lower in the fourth quartile compared to the first quartile ( $p < .001$ ) and the second quartile ( $p = .001$ ), but not compared to the third quartile ( $p = .08$ ). We once again calculated effect sizes; Cohen's *d* occupied a medium value when comparing the fourth and the third quartile ( $d = 0.53$ ) and high values when comparing the fourth and the second ( $d = 0.99$ ) and the fourth and the first quartile ( $d = 1.18$ ). Analyses related to absolute performance will now be followed by analyses related to relative performance.

Table 2 shows actual percentiles, predicted percentiles and differences between expected and actual percentiles for each quartile. The actual percentile increased gradually from the bottom to the top quartile, while the variable varied similarly around the mean within all quartiles. Values of the expected percentile also increased from the first to the fourth quartile, but this increase was far less even and steep compared to the actual percentile. Standard deviations also showed a greater discrepancy; variability was somewhat higher in the intermediate quartiles. Moreover, the calculated difference between the predicted and actual percentile had a positive valence in the first and the second quartile, meaning that these groups of participants overestimated their position in the sample; more specifically, position in the pattern was overestimated the most by the least skilled participants. Participants from the remaining two quartiles, on the other hand, underestimated their performance relative to others; the position in the sample was underestimated the most by participants from the top quartile.

Table 1. Actual and predicted absolute performance in points (quartiles)

	<i>N</i>	<i>M</i>	<i>SD</i>
Bottom quartile			
Actual absolute score	22	6.01	2.87
Predicted absolute score	22	13.43	4.33
Difference	22	7.42	4.79
Second quartile			
Actual absolute score	22	11.38	1.08
Predicted absolute score	22	18.12	5.50
Difference	22	6.74	5.28
Third quartile			
Actual absolute score	21	14.84	0.95
Predicted absolute score	21	19.28	5.07
Difference	21	4.43	5.28
Top quartile			
Actual absolute score	22	20.30	2.60
Predicted absolute score	22	22.04	4.10
Difference	22	1.73	4.82

Table 2. Actual and predicted relative performance (quartiles)

	<i>N</i>	<i>M</i>	<i>SD</i>
Bottom quartile			
Actual percentile	22	12.75	7.62
Predicted percentile	22	39.09	14.36
Difference	22	26.34	14.48
Second quartile			
Actual percentile	22	37.85	7.64
Predicted percentile	22	48.41	17.14
Difference	22	10.56	18.03
Third quartile			
Actual percentile	22	63.19	7.46
Predicted percentile	22	52.28	17.57
Difference	22	-10.92	19.33
Top quartile			
Actual percentile	22	88.51	7.42
Predicted percentile	22	55.90	13.77
Difference	22	-32.60	16.22

Results of a one-way ANOVA showed that the difference between the predicted and the actual percentile differed statistically significantly between groups,  $F(3, 84) = 49.48$ ,  $p < .001$ ,  $\eta_p^2 = 0.64$ . Post-hoc analyses showed that participants in the bottom quartile overestimated their position in the sample significantly more than did participants in the second ( $p = .003$ ,  $d = 0.97$ ), third ( $p < .001$ ,  $d = 2.18$ ) and top ( $p < .001$ ,  $d = 3.83$ ) quartile. Similarly, participants in the top quartile underestimated their position in the sample significantly more than participants in the third ( $p < .001$ ,  $d = 1.22$ ), second ( $p < .001$ ,  $d = 2.52$ ) and first ( $p < .001$ ,  $d = 3.83$ ) quartile.

The observed pattern is summarized in Figure 1. The graph on the left displays the difference between the expected and the actual absolute performance (for each quartile), while the graph on the right displays the difference between the expected and the actual relative performance (for each quartile).

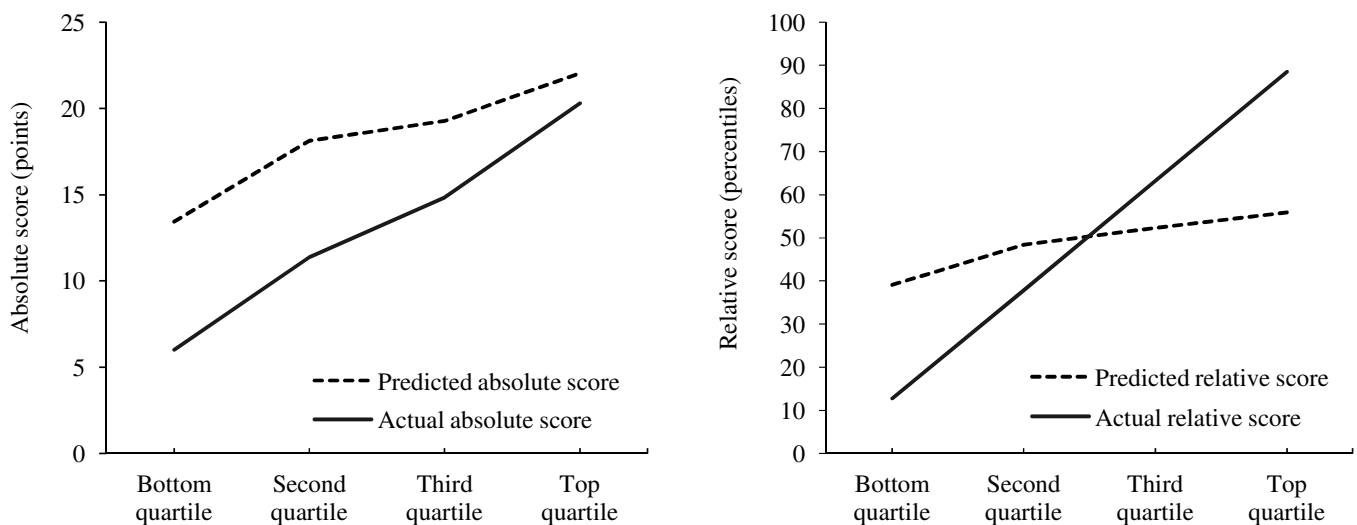


Figure 1. Differences between the actual and the predicted absolute score (left); the actual and the predicted relative score (right).

## Overclaiming

Concerning overclaiming, we first wanted to know how often participants claimed to be familiar with concepts (in our case literary works) that do not actually exist. Thirty-eight participants (43.7%) did not claim to be familiar with any fictitious concepts, while the remaining 49 participants (56.3%) claimed to know at least one fictitious concept. Of the latter, 34 participants claimed to be familiar with one fictitious work, 11 participants with two, four participants with three, and none with all four fictitious literary works by a Slovene author.

We were also interested in the relation between the self-perceived competence, the number of familiar real concepts and the number of familiar fictitious concepts. Self-perceived expertise correlated significantly with the number of familiar real concepts ( $r = .41$ ,  $p < .001$ ) and the number of familiar fictitious concepts ( $r = .36$ ,  $p = .001$ ); both correlation coefficients can be labelled as moderate. A significant correlation between the number of familiar real and fictitious concepts was also observed ( $r = .24$ ,  $p = .03$ ). The relation between self-perceived competence and the number of familiar real and fictitious concepts is illustrated in Figure 2.

In the overclaiming part of our research, half of the participants ( $N = 43$ ) assessed their own competence before tackling the overclaiming task, while the other half ( $N = 44$ ) assessed their competence after completing the overclaiming task. Before the main analysis, we checked whether the order manipulation influenced the assessment of competence. While the results implied that self-perceived competence was slightly higher when participants assessed it before completing the overclaiming task ( $M = 3.91$ ,  $SD = 1.23$ ) compared to when they assessed it after ( $M = 3.57$ ,  $SD = 1.11$ ), the difference was not statistically significant ( $U = 789.00$ ,  $Z = -1.38$ ,  $p = .17$ ,  $d = 0.29$ ).

Additionally, the comparison of the number of familiar fictitious concepts indicated that overclaiming was slightly higher in the group that assessed their competence before

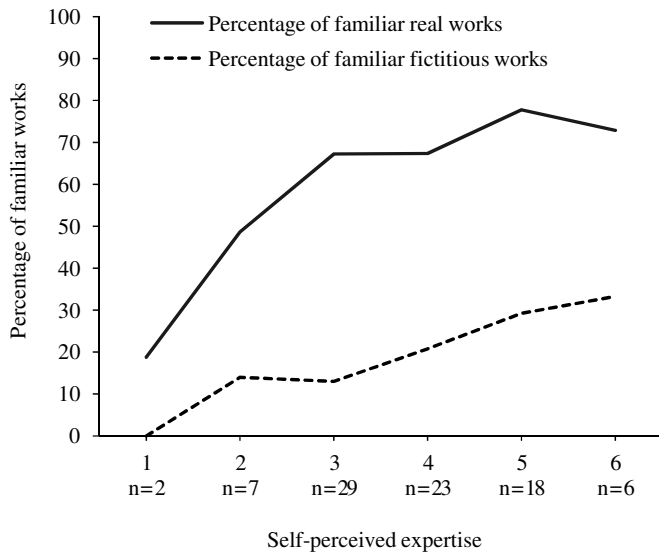


Figure 2. The relation between self-perceived competence and the percentage of familiar real and fictitious concepts.

completing the overclaiming task ( $M = 0.86$ ;  $SD = 0.83$ ), but the difference was relatively small (self-perceived competence after:  $M = 0.70$ ;  $SD = 0.85$ ). This was further illuminated by the Mann-Whitney  $U$  test, which showed that the difference between the two groups was not statistically significant ( $U = 832.50$ ,  $Z = -1.04$ ,  $p = .30$ ). The calculated effect size was low as well ( $d = 0.19$ ).

### Illusion of knowledge

In the last part of our study, participants were divided into two groups: a control group (low amount of information) and an experimental group (high amount of information). In both groups, participants were only slightly familiar with nanotechnology (control group:  $M = 2.11$ ,  $SD = 0.65$ , experimental group:  $M = 2.21$ ,  $SD = 0.81$ ) and the difference between them was not statistically significant ( $U = 911.00$ ,  $Z = -0.51$ ,  $p = .61$ ,  $d = 0.14$ ).

In the experimental group ( $N = 43$ ,  $M = 48.28$ ,  $SD = 16.80$ ), the average certainty in chosen answers was almost 10% higher than in the control group ( $N = 45$ ,  $M = 38.42$ ,  $SD = 15.16$ ). The variability was also slightly higher in the first, experimental group. An independent samples  $t$ -test showed that the difference between the groups was statistically signifi-

cant,  $t(86) = -2.89$ ,  $p = .005$ ,  $d = 0.62$ . This effect remained when controlling for differences in the initial familiarity with nanotechnology,  $F(1, 84) = 7.38$ ,  $p = .008$ ,  $\eta_p^2 = .08$ .

### Is overestimation a general or a domain-specific phenomenon?

At last, we checked whether people tend to overestimate their knowledge in different situations and domains. If this was true, absolute and relative overestimation in the Dunning-Kruger task (grammar), the extent of overclaiming (bibliography of a Slovene author), and certainty in wrong answers (nanotechnology) should all be strongly positively correlated. However, the results implied that this was not the case; neither in version 1 nor in version 2 of our instrument these variables correlated significantly (Table 3).

### Discussion

In the present study, our first goal was to find out whether people generally overestimate their knowledge. As predicted by *H1a*, participants indeed overestimated their absolute achievement. While our results thus largely replicate previous findings (e.g. Bell & Volckmann, 2011), a more detailed analysis interestingly shows that the overestimation observed in our study was more pronounced than in many earlier studies. We propose that this could be due to the high degree of difficulty of our test. Perhaps due to an implicit theory that the majority will score at least 50% (created on the basis of past experience, e.g. college exams), the average predicted absolute score was above this point (57%) and thus far above the actual absolute score which was approximately 40%. Alternatively, an unusually pronounced overestimation could be attributed to the use of the correction for guessing, which may have led to an even more distorted absolute self-assessment. In contrast to *H1a*, *H1b* focuses on relative overestimation. At first glance, our results do not support *H1b* and suggest a rather accurate relative self-perception, thus contradicting previous studies (e.g. Kruger & Dunning, 1999; Pavel et al., 2012). However, a more thorough analysis reveals that the small difference between the actual and the predicted percentile was largely due to the balance that suggests accurate self-perception, but contains both gross overestimation (bottom two quartiles) and substantial underestimation (top two quartiles). Additionally, the fact that the majority of participants did not place their relative performance above the average (as in Kruger & Dunning, 1999) could again be attributed to the high difficulty of our test; judging by their comments after testing, participants perceived the test as highly difficult and that could have been the reason behind more cautious judgments. Moreover, such results could be attributed to the characteristics of our specific sample – most of the participants were psychology students who had to have excelled on the Matura exams (an important part of which is also a grammar test) to get into the programme; it is hence possible that participants had the following mindset: “I did OK, but others performed equally or better”. Since we did not manipulate the difficulty of the test or the characteristics of the sample, these explanations require further testing.

Table 3. Coefficients of correlation between different tasks

	Version 1			Version 2		
	1	2	3	1	2	3
1. D-K: absolute overestimation						
2. D-K: relative overestimation	.75**			.71**		
3. Overclaiming	.19	-.04		-.03	.13	
4. False certainty	.07	.01	.11	.22	.19	.12

\* $p < .05$ , \*\* $p < .01$ .



The next two hypotheses were focused on the bottom quartile. We predicted that participants from the bottom quartile would overestimate their absolute performance the most. While the observed pattern of overestimation as well as the level of overestimation observed in the bottom quartile highly resemble previous studies (e.g. Bell & Volckmann, 2011), the differences observed in our sample cannot be confidently generalized; the bottom quartile did not overestimate their absolute knowledge statistically significantly more than the second and third quartile (though the effect size for the comparison between the bottom and the third quartile implies a medium effect). We argue that this finding, while interesting, does not necessarily speak against the core thesis that less competent participants give more inflated judgments. Since the process of dividing participants into quartiles was highly arbitrary, we could instead compare less-skilled (bottom half) and more-skilled (upper half) participants, which would result in a clearer conclusion that less-skilled participants overestimate their absolute achievement to a higher extent. Additionally, this discrepancy with past literature could partly be due to the somewhat low variability in absolute test scores – the test was not perfect and allowed only a fairly narrow range of scores. In contrast to *H2a*, our results clearly support *H2b*; participants from the bottom quartile overestimated their relative performance the most. Such a finding is consistent with previous studies (e.g. Pavel et al., 2012; Pazicni & Bauer, 2014).

Hypotheses *H3a* and *H3b* focus on the top quartile instead of the bottom one. While the participants from the top quartile indeed perceived their absolute knowledge more accurately than participants from the bottom and the second quartile, the difference between the third and the top quartile was not statistically significant. However, effect sizes do imply that – given a slightly larger sample – all comparisons between the top quartile and other quartiles would reach statistical significance. The observed pattern is therefore largely consistent with previous studies (e.g. Bell & Volckmann, 2011; Kruger & Dunning, 1999). We also predicted that the top quartile would underestimate their performance relative to others the most. Results obtained on our sample support this prediction. Such findings are mostly consistent with the existing body of literature (e.g. Kruger & Dunning, 1999; Pazicni & Bauer, 2014), though some studies showed a slightly lower discrepancy between the predicted and the actual percentile in the top quartile. In our opinion, our findings can be attributed to the combination of the false consensus effect (i.e. someone's overestimation of the extent to which their knowledge is normal; Ross, Greene, & House, 1977) and the specific attributes of our sample; it is possible that participants evaluated their classmates especially favourably, because of the high entry criteria they had to attain to get into the programme. Results related to *H3a* and *H3b* hence imply that the most-skilled participants knew that they had done a relatively good job on the test, but thought that other participants had been similarly or more successful.

In sum, results from the first part of our study show that – though there are some small deviations, which could be a result of a thorough and highly difficult test – the Dunning-Kruger effect does not look vastly different when knowledge

is assessed with many different types of tasks and when the task at hand is highly difficult. More specifically, despite a highly difficult test, poor performers still grossly overestimated their absolute and relative performance, showing a level of miscalibration that is largely inconsistent with claims by Burson et al. (2006).

We now move to the second phenomenon – overclaiming. As predicted by *H4*, the majority of respondents claimed knowledge of at least one fictitious concept. However, the share of those who overclaimed was much lower than in previous studies (e.g. Atir et al., 2015). We propose two explanations for this discrepancy. First, this might be partly due to a specific topic (i.e. a bibliography of Ivan Cankar) as opposed to a broader topic (i.e. biology or literature). Second, we believe that lower proportion of overclaiming could be attributed to a different response format – past studies measured overclaiming almost exclusively with a 7-point Likert scale, while we decided to measure it dichotomously. We claim that a dichotomous response format could be both less misleading as well as a better reflection of reality. In sum, these results show that overclaiming, even when measured dichotomously, is common, but perhaps not as prevalent as shown by previous studies.

Additionally, our results support *H5a*; we found a positive relationship between self-perceived expertise and the number of familiar real concepts. This result is consistent with previous literature (Atir et al., 2015) and implies that self-perceived expertise is not completely distorted. Our results also support *H5b*, which predicted that there would be a positive relation between self-perceived expertise and the extent of overclaiming. Such a finding is a successful replication of the study by Atir and colleagues (2015) and strongly implies that people make judgments about what they know based on their perception of knowledge in a certain domain. *H6* tested the order effect; while overclaiming was a bit more pronounced when participants assessed their competence before responding to the main task, the difference was not significant. As it stands, it does not matter whether participants consciously think about their expertise (and write their answer down) before the main overclaiming task; participants' perceived expertise is something that affects answering in all situations. Hence, our results are consistent with the results reported in the first study by Atir et al. (2015), but contradict results from the second part of the same study.

Our last hypothesis, *H7*, was related to the illusion of knowledge; we predicted that participants who received more information about nanotechnology, would be more certain in their answers. The results obtained on our sample clearly speak in favour of this hypothesis. Such a finding is consistent with previous theories and studies conducted in the United States (e.g. Gill et al., 1998; Heath & Tversky, 1991; Schwartz, 2004) and implies that an increase in quantity (but not quality) of information can result in higher certainty. Lastly, we also calculated correlations between absolute and relative overestimation in the field of grammar, the extent of overclaiming when judging familiarity with the works of Ivan Cankar, and certainty in wrong answers about nanotechnology, and found only low correlations between the measures. This finding contributes to the growing body of literature which recog-

nizes overconfidence as a domain-specific trait (e.g. Kruger & Dunning, 1999), but further illuminates how very nuanced these metacognitive judgments really are; for example, grammar and history of Slovene literature – domains that are normally seen as closely related – lead to widely disparate judgments. Additionally, as we did not manipulate domains in isolation but rather vary the domains and types of tasks at the same time, the lack of significant correlations also supports the notion that phenomena included are largely independent and not just elements of a general and stable trait.

## Limitations and conclusions

Many segments of our instrument included alterations to the well-established ways of measuring these phenomena; while these modifications can be understood as valuable considerations about improvements needed in this area of research, they can also represent key shortcomings of our study, especially regarding comparison with previous studies. Additionally, as we did not, for example, manipulate the difficulty of the grammar test or the response format in the overclaiming task, we cannot talk about causes and effects; hence, the present study only describes what happens in altered conditions without a clear comparison with well-established ways of measuring these phenomena. Further studies should therefore test these ideas in a systematic program of research.

Despite these limitations, our study illuminates various deficiencies in self-assessment. Only by collecting and verifying information about the various lacunae in the perception of our own knowledge can we take the right steps towards improvement – towards achieving more accurate self-assessment, which could, in the next step, as indicated by our introductory examples, improve our society.

## References

- Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science, 26*, 1295–1303.
- Bell, P., & Volckmann, D. (2011). Knowledge surveys in general chemistry: Confidence, overconfidence, and performance. *Journal of Chemical Education, 88*, 1469–1476.
- Bishop, G. F., Oldendick, R. W., Tuchfarber, A. J., & Bennett, S. E. (1980). Pseudo-opinions on public affairs. *Public Opinion Quarterly, 44*, 198–209.
- Bradley, J. V. (1981). Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society, 17*, 82–84.
- Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*, 60–77.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.
- Dunning, D. (2014, October 27). We are all confident idiots. *Pacific Standard*. Retrieved from <http://www.psmag.com/health-and-behavior/confident-idiots-92793>
- Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*, 1082–1090.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5–17.
- Everson, H. T., & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science, 26*, 65–79.
- Gill, M. J., Swann, W. B. Jr., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75*, 1101–1114.
- Golčič, J. (1972). Razumevanje tujjica, verbalizam i semantički snobizam [Understanding of foreign words, verbalism and semantic snobbism]. V *Psihološke razprave: IV. kongres psihologov SFRJ, Bled, 13.-17. X. 1971* [Psychological debates: IV. Congress of SFRJ psychologists, Bled, 13.-17. X. 1971] (pp. 69–72). Ljubljana, Slovenia: Društvo psihologov Slovenije in Filozofska fakulteta v Ljubljani.
- Gregersen, N. P. (1994). Systematic cooperation between driving schools and parents in driver education, an experiment. *Accident Analysis & Prevention, 26*, 453–461.
- Gregersen, N. P. (1996). Young drivers' overestimation of their own skill – an experiment on the relation between training strategy and skill. *Accident Analysis & Prevention, 28*, 243–250.
- Gross, M., & Latham, D. (2012). What's skill got to do with it? Information literacy skills and self-views of ability among first-year college students. *Journal of the American Society for Information Science and Technology, 63*, 574–583.
- Hall, C. C., Ariss, L., & Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes, 103*, 277–290.
- Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty, 4*, 5–28.
- Kahan, D. M., Braman, D., Slovic, P., Gastil, J., & Cohen, G. (2009). Cultural cognition of the risks and benefits of nanotechnology. *Nature Nanotechnology, 4*, 87–90.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.
- Kunkel, E. (1983). Driver improvement courses for drinking-drivers reconsidered. *Accident Analysis & Prevention, 15*, 429–439.
- McKenna, F. P., Stanier, R. A., & Lewis, C. (1991). Factors underlying illusory self-assessment of driving skill in males and females. *Accident Analysis & Prevention, 23*, 45–52.

- Paulhus, D. L., & Harms, P. D. (2004). Measuring cognitive ability with the overclaiming technique. *Intelligence, 32*, 297–314.
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 890–904.
- Pavel, S. R., Robertson, M. F., & Harrison, B. T. (2012). The Dunning-Kruger effect and SIUC University's aviation students. *Journal of Aviation Technology and Engineering, 2*, 125–129.
- Pazicni, S., & Bauer, C. F. (2014). Characterizing illusions of competence in introductory chemistry students. *Chemistry Education Research and Practice, 15*, 24–34.
- Public Policy Polling (2015). *National survey results*. Retrieved from <http://www.publicpolicypolling.com/pdf/2015/GOPResults.pdf>
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 4*, 332–348.
- Sheldon, O. J., Dunning, D., & Ames, D. R. (2014). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology, 99*, 125–137.
- Swami, V., Papanicolaou, A., & Furnham, A. (2011). Examining mental health literacy and its correlates using the overclaiming technique. *British Journal of Psychology, 102*, 662–675.
- Tracey, J., Arroll, B., Barham, P., & Richmond, D. (1997). The validity of general practitioners' self assessment of knowledge: Cross sectional study. *British Journal of Medicine, 315*, 1426–1428.