

Running Head: HUMOR ABILITY

Ha Ha? Assessing Individual Differences in Humor Production Ability

Emily C. Nusbaum

Yale University

Paul J. Silvia

Roger E. Beaty

University of North Carolina at Greensboro

Author Note

Emily C. Nusbaum, Yale Center for Emotional Intelligence, Yale University; Paul J. Silvia and Roger E. Beaty, Department of Psychology, University of North Carolina at Greensboro.

The data presented in this paper has been de-identified and posted online at Open Science Framework <https://osf.io/8q6b9/>.

Correspondence should be addressed to Emily C. Nusbaum at the Yale Center for Emotional Intelligence, P.O. Box 208376, New Haven, CT, 06520-8376, or to Paul J. Silvia at the Department of Psychology, P.O. Box 26170, UNCG, Greensboro, NC, 27402-6170; or by e-mail at emily.nusbaum@yale.edu or p_silvia@uncg.edu.

Abstract

Humor is one of the most salient examples of verbal creativity in everyday life, but relatively little is known about individual differences in the ability to be funny. The present research examines the assessment of humor production ability—the ability to generate funny ideas on the spot. With only a few exceptions, humor ability research has relied on a single task: asking people to write captions for single-panel cartoons. In three studies, we evaluate the cartoon-captions task alongside a recent résumé-completion task developed by Howrigan and MacDonald (2008) and two new tasks that we developed: a joke-completion task (writing a funny conclusion to a joke set up by the researcher) and a definitions task (writing a funny definition for an odd noun-noun pair, such as *yoga bank* and *cereal bus*). In all three studies, the newer tasks covaried strongly with the cartoon captions task, suggesting that they measure the same underlying humor production ability. Of the major personality factors, measured based on the NEO (Studies 1 and 2) and HEXACO (Study 3) models, only openness to experience significantly predicted humor ability, and its effects were medium and large in size ($\beta = .48$ [.34, .63]; $\beta = .54$ [.39, .70]; $\beta = .36$ [.22, .50]). The findings suggest that humor ability shares much in common with other forms of verbal creativity, and that researchers could adopt a multimethod approach to measuring it.

Keywords: humor, creativity, personality, openness to experience, assessment

Ha Ha? Assessing Individual Differences in Humor Production Ability

Humor is important to everyday life. For years researchers have reported relationships between humor and social, emotional, and physical well-being (Boyle & Joss-Reid, 2004; Celso, Ebener, & Burkhead, 2003; Fry, 1994; Kuiper & Nicholl, 2004; Martin, 2004). In particular, laughing has been associated with reducing the impact of everyday stressors on health (Svebak, 2005), and humor is linked to social functioning in both close relationships and superficial interactions (Bressler & Balshine, 2006; Dyck & Holtzman, 2013; Polimeni, Campbell, Gill, Sawatzky, & Reiss, 2010; Sprecher, & Regan, 2002; Stanley, Lohani, & Isaacowitz, 2014).

Humor ability—the ability to generate funny ideas on the spot—is probably the side of humor that is most relevant to creativity research. Individual differences in humor ability are huge: some people are hilarious, others are painfully unfunny, and most are somewhere in between. Why people differ in humor ability has not been widely studied, but it is increasingly attracting attention from researchers interested in creativity, cognitive abilities, and evolutionary psychology (Bressler, Martin, & Balshine, 2006; Greengross, Martin, & Miller, 2012; Greengross & Miller, 2011; Kaufman, Kozbelt, Bromley, & Miller, 2008; Kozbelt & Nishioka, 2010).

Like many abilities involving creativity, humor ability raises complex measurement issues. Most studies of humor production employ the same assessment: the cartoon caption task. In this task, people are presented a cartoon (usually a single-panel cartoon) and are asked generate a funny caption for it. Raters then score the captions for funniness. This popular task has many variations. Some studies ask people to write multiple captions for one cartoon, whereas others limit people to writing one really good caption. Likewise, some studies impose time limits, whereas others allow unlimited time. Most studies ask people to come up with their own funny captions, whereas a few force people to complete fill-in-the-blank, partially-composed captions. And finally, the number of cartoon items, the actual cartoons used, the number of raters, and the rubric for rating the participants' ideas vary widely.

The cartoon caption task has dominated past humor research. Some tasks and scales do earn “gold standard” status in fields of research, but it is usually a bad sign when a single task is used in nearly all of a field’s studies. A more diversified set of assessment tools would reduce the influence of task-specific quirks on a literature and afford statistical models that pool across methods. The cartoon captions task has served humor research well, but additional complementary tasks would move this growing field forward. Just as the study of creative cognition should not rely exclusively on unusual uses tasks, the study of humor production should develop a portfolio of practical tasks that capture different sides of the underlying ability.

The present study thus examines classic, recent, and brand new humor production tasks. Our goal was to evaluate the classic cartoon captions task in light of some promising newcomers. Some of the new tasks have been proposed in recent research, such as a clever résumé writing task developed by Howrigan and MacDonald (2008). Others are new tasks that we developed, such as completing open-ended jokes and creating funny definitions for eccentric noun pairs.

To gain evidence for validity, we explored the relationship between humor production and the major factors of personality—specifically, the Big Five traits and the closely related HEXACO model (Lee & Ashton, 2004). Research on humor ability and personality is sparse (Nusbaum & Silvia, in press), but two traits have attracted the most attention. The first, openness to experience, is not surprising, given the prominence of openness to experience in creativity (Fürst, Ghisletta, & Lubart, in press; Kaufman, 2013; Oleynick et al., in press). Humor ought to be closely linked to openness to experience, given that it involves generating new ideas, applying verbal abilities, and thinking unconventionally (Earleywine, 2010). Several studies (but not all; Greengross, Martin, & Miller, 2011 [professional comedian subsample]; Moran, Rain, Paige-Gould, & Mar, 2014) have found significant relationships between openness to experience and scores on the cartoon caption task. The second, extraversion, has a long history in humor research. Extraversion is associated with being outgoing, gregarious, cheerful, and

light-hearted (Watson, 2000; Wilt & Revelle, 2009). Some studies have found that extraverted people generate funnier ideas, but these findings are inconsistent and controversial (Babad, 1974; Howrigan & MacDonald, 2008; Köhler & Ruch, 1996).

In the present set of studies, we gave several humor production tasks to three samples of college undergraduates. All three studies included the cartoon captions task alongside the newer tasks. Personality was measured in each study to provide evidence for the validity of the tasks' scores. As a group, the studies illuminate the performance of new methods of measuring humor ability and illustrate the personality traits that typify funny people.

Study 1

Study 1 explored the cartoon captions task alongside two newcomers: a résumé completion task (Howrigan & MacDonald, 2008) and a joke completion task. People completed all three tasks, which were examined in relation to each other and to measures of the Big Five personality factors.

Method

Participants and Design

A final sample of 166 students at the University of North Carolina at Greensboro (UNCG) participated in this study. This sample was comprised of mostly Caucasian (49%), African American (31%), and Asian (12%) women (77%). Typical of university samples, the average age of participants was 19 years ($SD = 2.9$). The humor tasks were completed in the same order for all participants (résumés first, captions second, jokes third) and were followed by the personality surveys. For each task, the order of the items was randomized.

Procedure

People participated in groups ranging from 1 to 8. Upon arriving at the lab room, participants received a consent form to read and sign. The experimenter instructed people briefly on the three different humor tasks before beginning the assessments. All tasks were completed on computers using

MediaLab (Empirisoft, NY), except the résumé tasks, which were completed on paper handouts.

Humor assessment. To evaluate humor production ability, people completed three different humor tasks, each of which had three items. All participants completed each item for each task, for a total of 9 items per participant. The *captions task* asked people to write a funny caption for each of three caption-less *New Yorker* cartoons. This task is by far the most common task in humor assessment (Nusbaum & Silvia, in press), although the actual cartoons presented to participants vary across studies (Babad, 1974; Feingold, 1983; Feingold & Mazzella, 1993; Koppel & Sechrest, 1970; Turner, 1980). For this 6-minute task, people were instructed to simply write a funny caption for the blank cartoons; they saw an example before beginning the task. One of the three cartoons depicts an astronaut on the moon talking on a cell phone, another shows a man in a crown lying on a sofa talking to a psychotherapist, and the third shows two business men, one with a smoking gun, standing next to a corpse on the floor of an office.

A second task assessing humor production, the *joke stems* task, provided people with a funny scenario and the beginning of a joke that people could finish by writing something funny. People had 6 minutes to complete three different joke stems, which were given to participants in randomized order. The joke tasks were presented as a set-up and a task. One task said:

Imagine that one of your friends wants your opinion on how well she sings. She sings a minute or two to demonstrate her voice, and you cringe—she might be the worst singer you’ve ever heard. When she asks, “So how was it?” you decide to be honest, so you say, “To be honest, listening to that was like...”

People were then asked to complete the phrase “To be honest, listening to that was like...” with something funny. All three joke stems are described in the Appendix. Researchers have used joke completion tasks in the past, but their tasks have typically taken the form of a multiple choice test, a joke recognition test, or joke comprehension task (i.e., Feingold & Mazzella, 1991, 1993). Our new task is

thus more clearly focused on producing ideas that are novel and funny instead of recognizing or selecting funny ideas created by others.

Our final humor production task, the *résumés task*, asked participants to complete a blank résumé for each of three given photos of strangers. For this task, people were given the following instructions and explicitly told to be funny:

The following task presents you with blank résumés accompanied by head shots. You should examine the photos that are presented, and imagine some characteristics of each subject. Fill in the blank résumé form that accompanies each photo.

As part of the résumé, people completed the following for each photo: *Name, Occupation, About Me, Hobbies/Activities, My Typical Day, and Philosophy of Life*. The three pictures were of a young man with unusually voluminous hair staring into the distance, a young woman in goth-style dark hair with heavy, dark make-up, and a photo of a swimming dog from Seth Casteel's (2012) photography book *Underwater Dogs*. The résumé task was developed by Howrigan and MacDonald (2008), who found good evidence for its reliability (6 résumés with four raters; Cronbach's $\alpha = .72$). Participants were allowed to work on this task at their own pace and for as long as they liked.

Just as divergent thinking tasks should tell participants to "be creative" so that they understand the task goal (Nusbaum, Silvia, & Beaty, 2014), humor tasks should emphasize that participants should "be funny." The experimenter's instructions at the start of the study and the written instructions in MediaLab emphasized that the study was about how people come up with funny ideas, that the humor tasks were intended to measure funniness, and that people should try to come up with funny ideas that would make other people laugh.

Our approach to scoring the humor tasks grew out of our past research in scoring open-ended creativity tasks, such as divergent thinking (Silvia et al., 2008) and metaphor production (Beaty & Silvia, 2013). Each response for the 9 different humor items (3 joke stems, 3 captions, 3 résumés) was scored

by 4 raters using subjective scoring methods. The 4 raters rated each response on a 1 (*not at all funny*) to 5 (*very funny*) scale. Though the ratings were given as a holistic judgment for each response, raters were given instructions (inspired by Wilson, Guilford, and Christensen's (1953) definition of creative ideas, which we have used before; Silvia et al., 2008) to give higher ratings to responses that struck them as uncommon or unusual, funny, and clever, and lower ratings to responses that they found irrelevant, unsuitable, or boring.

The responses for each computerized task were given a random ID and then sorted alphabetically so that no identifying information accompanied the responses as they were rated. For the résumé handouts, a random subject ID was written on each page of the handout, and the pages were separated, shuffled, and sorted by picture. All raters gave their scores without knowledge of the other raters' scores or any of the participants' other scores.

Personality assessment. To explore individual differences that are related to humor production ability, we measured personality using the NEO-FFI 3 (McCrae & Costa, 2007), a 60-item version of the NEO Personality Inventory. The NEO-FFI 3 measures the five broad factors of personality—neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness (McCrae & Costa, 2008)—with 12 items per factor. People use a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*).

Results and Discussion

Data Reduction and Reliability Estimation

The scores for the humor tasks have a faceted structure: there are three humor tasks, each with three items rated by four raters. Based on recent recommendations by Primi (2014), we used many-facet Rasch models to evaluate reliability and distill the faceted data into a single score. The many-facet Rasch approach is an extension of conventional Rasch models. A conventional Rasch model for 5-point rating-scale data would estimate the probability that someone receives a particular scale score as a function of (1) the participant's latent trait level (humor ability, in this case), (2) the item's difficulty

(how hard or easy it is to get high scores on one of the items), and (3) a parameter representing the thresholds between the response options in the 1-5 scale. A many-facet approach adds a rater severity parameter, which represents how tough or lenient a rater was when judging humor. The virtues of this approach are many (Barbot, Tan, Randi, Santo-Donato, & Grigorenko, 2012; Primi, 2014). Researchers can estimate how severe each rater is and adjust each participant's estimated trait score in light of the severity of the raters and the difficulty of the items (Eckes, 2011).

The many-facet Rasch models were estimated using FACETS 3.71.4 (Linacre, 2014), using joint maximum likelihood estimation. We estimated three models, one for each task type (captions, joke stems, and résumés). For each model, participants, items, and raters were defined as facets. Cases with extreme infit or outfit values were dropped (1 person for the cartoons task, and 1 for the joke stems task), and the resulting estimated "fair average scores" (scores corrected for the difficulty of the items and severity of the raters) were saved for analysis in subsequent structural equation models. These fair average scores are on the same 1-5 scale as the raw scores.

Rasch person reliability was .51 for captions, .63 for joke stems, and .67 for résumés. Although these values are on the same 0 to 1 scale as conventional reliability metrics (e.g., Cronbach's alpha), they have a different interpretation. Rasch person reliability figures represent the lower-bound of the true reliability value (Eckes, 2011; Linacre, 1997), so they will be lower than Cronbach's alpha, which is inflated by extreme scores.

The 12 items for each of the five NEO factors were averaged to create summary scores. Humor ability was modeled as a latent variable with three indicators: the fair-average scores for each task derived from the many-facet Rasch models. Table 1 displays descriptive statistics and correlations. The classic captions task correlated well with the newer résumés task ($r = .37$) and jokes tasks ($r = .41$), and the jokes and résumés tasks correlated well with each other ($r = .44$).

Personality and Humor Ability

Who was funny? A structural equation model estimated the effects of the five personality factors on humor ability. The model was estimated in Mplus 7.3 using maximum likelihood with robust standard errors. Figure 1 depicts the model; the standardized coefficients and confidence intervals are displayed in Table 2. The fit of the model was very good on most fit indices: $\chi^2(df = 10) = 9.736, p = .464$; $\chi^2/df = .97$, CFI = 1.00, SRMR = .025, RMSEA = .000, 90% CI [.000, .083].

As expected, openness to experience had the largest relationship with humor ability ($\beta = .48, p < .001$) and a large effect size, using the .10/.30/.50 guidelines for effect sizes in the r metric (Cumming, 2012). People higher in openness to experience thus gave responses that were scored as much funnier by the raters. The remaining effects were small in size and not significant (see Table 2 for the effects and their confidence intervals). Notably, extraversion—a prominent trait in past humor research (e.g., Feingold & Mazzella, 1993; Köhler & Ruch, 1996)—had essentially no effect ($\beta = -.09, p = .443$).

Study 2

Study 2 sought to replicate the findings from Study 1, particularly for the new joke stems task that we developed. In Study 2, we thus again measured the Big Five personality traits and humor ability, which was assessed using the joke stems and captions tasks.

Method

Participants and Design

A sample of 147 college students participated in this study. The sample was typical of our university and was predominantly female (72%), Caucasian (72%), and African American (26%). On average, participants in this sample were 19 years old ($SD = 1.8$ years). This study was much the same as Study 1: the same three joke tasks and the same three caption tasks were used to assess humor ability in a between-person design, and we again measured personality with the NEO FFI 3.

Procedure

People participated in groups of 1 to 8 people. Upon entering the lab, participants were seated

at a computer and completed an informed consent form. People were then given general instructions on how to complete the experiment, and more specific instructions about how to complete the two different humor tasks. The entire experiment was completed on computers using MediaLab (Empirisoft, NY).

Humor assessment. This study assessed humor production ability using the same jokes and captions tasks used in Study 1. We gave participants the same instructions, and they wrote captions for the same cartoons and completed jokes for the same joke stems used in Study 1. Each of the six humor assessment items was scored by 5 independent raters using the same guidelines and the same scale as described in Study 1.

Personality assessment. In addition to replicating the humor assessments, we again included the NEO FFI 3 personality scale. Because we found such a large effect of openness in Study 1, we decided to use the same measure of personality to examine whether the effect would replicate in a new sample.

Results and Discussion

Data Reduction and Model Specification

As in Study 1, we used many-facet Rasch models to generate humor scores for each participant. We ran one model for the jokes task and another for the captions task. Each model defined participants, items, and raters as facets, and participants with extreme infit and outfit values were omitted (1 for jokes, 0 for captions). Rasch reliability was .60 for the jokes task and .62 the captions task. The estimated “fair average” scores were saved for use in later analyses. The descriptive statistics are shown in Table 3. As in Study 1, the newer jokes task correlated well with the classic captions task ($r = .42$).

Personality and Humor Ability

How did personality predict humor ability? Figure 2 depicts the model evaluating the effects of the five NEO factors on humor ability, a latent variable composed of Rasch scores for the jokes and

captions tasks. The fit of the model was very good on most fit indices: $\chi^2(df = 5) = 8.013, p = .156; \chi^2/df = 1.60, CFI = .954, SRMR = .031, RMSEA = .064, 90\% CI [.000, .142]$. The effects and confidence intervals, displayed in Table 2, closely resembled the effects from Study 1. As before, openness to experience had the largest effect and a notably large effect size ($\beta = .54, p < .001$). No other factor had a significant effect. Extraversion, as in Study 1, had essentially no relationship with humor ability ($\beta = -.04, p = .688$).

The findings from Study 1 and Study 2 thus converge on a few points. Openness to experience, by far, appears to be the most important factor for predicting people's ability to be funny on the spot. Furthermore, alternatives to the cartoon captions task, including the new joke completion task that we developed, apparently work well and thus expand the range of assessment options.

Study 3

Study 3 had two goals. First, we wanted to explore a new type of humor task: creating funny definitions for odd noun-noun combinations. In this task, participants were given a noun-noun combination (e.g., *yoga bank, cereal bus, balloon bench*) and asked to come up with a funny definition for it. We took our inspiration from the literature on the cognitive psychology of conceptual combinations (Costello & Keane, 2000; Wisniewski, 1997) and from the long comedic tradition of making definitions for neologisms and portmanteaus.

Second, we wanted to dig deeper into the relationships between personality and humor production ability. Studies 1 and 2 clearly demonstrated that of the Big 5 traits, openness to experience has the strongest relationship with humor production ability. A natural next step is to explore these links at the facet level. To do so, we assessed personality based on the six-factor HEXACO model (Lee & Ashton, 2004), which is similar to the traditional five-factor models in most regards but diverges by extracting a sixth factor (honesty-humility). The HEXACO's facet structure for openness to experience is intuitive and appealing—it specifies *creativity* (values creativity, has good ideas), *unconventionality* (quirky, off-beat, untraditional), *inquisitiveness* (curious, loves learning), and *aesthetic appreciation*

(enjoys literature and the fine arts) as the four facets—so it is a good model for looking at the effects of openness to experience at a finer level of detail.

Method

Participants and Design

A sample of 138 college students participated in this study. The sample was again quite typical of our university: 67% women, 59% Caucasian, and 33% African American. On average, participants in this sample were 18.7 years old ($SD = 1.9$ years). Like in Studies 1 and 2, this study assessed individual differences in humor production ability as they relate to personality. Participants completed task types (i.e., joke task, caption task, definition task) in the same order, but the individual items were randomized within task type.

Procedure

People participated in groups of 1 to 8. Participants completed consent forms upon entering the lab and were seated at a computer. People were then given general instructions on how to complete the experiment, and more specific instructions about how to complete the three different humor tasks. The entire experiment was completed on computers using MediaLab (Empirisoft, NY). The humor tasks were untimed because almost no one hit the time limits in the prior studies.

Humor assessment. This study assessed humor production ability using the exact same jokes and captions tasks that were used in Studies 1 and 2, and as before, we encouraged people to be funny. We elaborated somewhat more in the written MediaLab instructions for this study. For example, for the joke stems task, the instructions noted: “Your job is to finish the joke by writing something funny. What you write can be weird, silly, gross, bizarre, ironic, dirty-minded, or whatever, so long as it’s funny.” The purpose was to make participants comfortable with expressing bizarre or off-color responses, just in case some participants were holding back funny ideas.

In addition to those tasks, we introduced a new type of humor assessment that we call the

definitions task. For this task, we took some inspiration from the literature on conceptual combination (Costello & Keane, 2000; Wisniewski, 1997), much of which examines how people creatively construct meanings for novel pairs of concepts. Participants were shown an odd noun-noun combination and asked to come up with a funny definition for that new word. Specifically, participants were told the following:

A classic form of humor is coming up with funny definitions for things. So, for this next task, you will be given an unusual noun and asked to come up with a funny definition for it, something that most people would find funny or silly. It's fine to be weird, silly, dirty, ironic, bizarre, or whatever, so long as the definition is funny. For example, you might define "professor" as "someone who talks in someone else's sleep."

Following from the jokes and captions tasks, participants wrote one definition for each of three word pairs: *yoga bank*, *snuggle war*, and *fruit jar*. Two independent raters scored each of the nine humor assessment tasks using the same guidelines and the same scale as described in Study 1.

Personality assessment. Participants completed the 100-item HEXACO-PI-R (Lee & Ashton, 2004). This test assesses the six HEXACO traits (Honesty-humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to experience). The HEXACO-100 gives six global factor scores as well as four facet scores per factor. For openness to experience, the four facets are *creativity*, *aesthetic appreciation*, *inquisitiveness*, and *unconventionality*. The HEXACO model and the HEXACO-PI-R have been examined across a wide variety of samples (Ashton & Lee, 2009; Ashton, Lee, Perugini, Szarota, de Vries, et al., 2004). Each factor is measured with 16 items (the 4 remaining items assess altruism, an interstitial factor) that use a 5-point response scale (1 = *strongly disagree*, 5 = *strongly agree*).

Results and Discussion

Data Reduction and Model Specification

A humor ability score was estimated using many-facet Rasch models: participants, raters, items, and tasks were defined as facets. Unlike the prior studies, a single model—rather than separate models for each humor task—was estimated because the smaller number of raters (2) makes the data too sparse for task-specific facet analyses. The Rasch reliability of the resulting fair-average scores was .56.

As before, the three humor tasks correlated well with each other (see Table 4). When the raters' scores were averaged, the captions task correlated strongly with the joke completion task ($r = .51$) and the definitions task ($r = .45$), and the joke and definitions tasks correlated strongly with each other ($r = .50$).

Personality and Humor Ability

How did personality predict humor ability? Figure 3 depicts the model evaluating the effects of the six HEXACO factors on humor ability, which in this case is an observed score. The effects, shown in Table 2, again resemble the effects from Studies 1 and 2. Openness to experience had the largest effect size ($\beta = .36, p < .001$) and the only significant effect. Notably, the effect size was medium in size and thus smaller than the effects for NEO openness found in Studies 1 and 2. Extraversion, as before, had essentially no relationship with humor ability ($\beta = -.04, p = .606$).

We next explored the four HEXACO facets of openness to experience to see if any particularly stood out as a predictor of humor ability. The simple correlations between the humor score and the facets were estimated. Humor correlated the highest with unconventionality ($r = .36 [.21, .51], p < .001$) and aesthetic appreciation ($r = .34 [.19, .48], p < .001$), more modestly with inquisitiveness ($r = .24 [.06, .41], p = .009$), and weakly with creativity ($r = .13 [-.03, .28], p = .106$). No obvious message emerges from the pattern of facet correlations. On the one hand, unconventionality should be important to humor, given models of humor that emphasize the importance of violating a mental model of what the world is like (Earleywine, 2010). On the other hand, one would probably expect the creativity facet—viewing oneself as a creative person and having creative interests—to be important to humor, but it had

the smallest effect. The facet findings thus await replication and future research.

Discussion

What are funny people like? Answering this question requires a method for measuring the ability to be funny, an ability that has not attracted much psychometric attention. Since the early work of Koppel and Sechrest (1970), nearly all research has relied on some form of a cartoon captions task to measure humor ability. In the present research, we evaluated a group of newer tasks, including two new tasks presented for the first time.

All the newer tasks—résumés, joke completion, and definitions—worked well. Their reliability estimates were at least as good as the estimates for the cartoon captions task, and they correlated well with the captions task. The correlations between the tasks were nearly the same, so the new tasks correlated as highly with each other as they did with the captions task. Good evidence for the validity of the tasks' scores comes from their relationships with openness to experience. All three studies found effect sizes that were medium to large in size, a finding that replicated across different raters, personality inventories, and groups of tasks. The substantial effects for openness to experience are consistent with the broader literature on how personality predicts creative thought, and they should increase researchers' confidence in the validity of the tasks' scores.

Regarding reliability, the Rasch reliability estimates reported here aren't easily compared to reliability estimates from past research. First, rating data creates faceted rater \times item designs, but traditional reliability metrics evaluate one facet while ignoring the other (Eckes, 2011). For example, Cronbach's alpha will estimate score reliability for the items, pooled across raters, and metrics like kappa will estimate reliability of the raters, pooling across items. Holistic estimates of reliability for the faceted design—such as those provided by generalizability theory (Brennan, 2001; Cronbach, Nageswari, & Gleser, 1963) or many-facet Rasch models—have not been reported in past work. Second, Rasch reliability estimates represent a floor, the lower-limit of the true reliability level, whereas Cronbach's

alpha is biased upward (Linacre, 1997). The Rasch reliability estimates in the three studies ranged from .51 to .67; a typical Cronbach's alpha for a cartoon captions task is around .60 to .75 (Feingold & Mazzella 1991; Feingold & Mazzella 1993; Greengross et al., 2012; Kozbelt & Nishioka, 2010). The medium and large effect sizes indicate that the tasks are performing at least reasonably well, but the reliability of humor tasks strikes us as somewhat lower than similar measures of creative thought (e.g., divergent thinking).

Personality and Humor

In all three studies, openness to experience had the largest effect size, and in two of them the effect size was around .50, a value seen as large overall (Cumming, 2012) and as a notably large effect size in creativity research. What is it about openness to experience that makes it such a strong predictor of humor ability? Openness to experience is a broad trait: people high in openness to experience view themselves as creative people (Karwowski & Lebuda, 2016); spend more time in creative pursuits (Conner & Silvia, 2015; Silvia et al., 2014); have higher fluid and crystallized intelligence (DeYoung, Peterson, & Higgins, 2005; DeYoung, Quilty, Peterson, & Gray, 2014; Ziegler, Danay, Heene, Asendorpf, & Bühner, 2012); appreciate music, literature, and the arts (Silvia, 2007; Swami & Furnham, 2014); have a broader emotional range (Terracciano, McCrae, Hagemann, & Costa, 2003) and experience states like awe, wonder, and chills more easily (Nusbaum et al., 2014; Nusbaum & Silvia, 2014); have a more vivid inner fantasy life (Zingrone, Alvarado, & Dalton, 1998); and value being unique and unconventional (Joy, 2004).

There is thus a lot going on with openness to experience. An interesting problem for future work is to untangle the many likely pathways from openness to experience to humor. One pathway likely involves abilities: people high in openness to experience have higher vocabularies, verbal abilities, and fluid abilities, all of which should help in crafting witty verbal material. Another path likely involves motivation: people high in openness should appreciate the essentially unconventional character of

humor, which necessarily pokes and tweaks an audience's concepts and sense of what the world is like (Dagnes, 2012; Earleywine, 2010). Digging into this relationship is a fertile direction for future research on humor and personality.

There is some inconsistency among other traits that deserves attention, particularly for extraversion. Because social activities are the domain of highly extraverted people, and because humor is an inherently social thing, it is natural to expect that people who are high in extraversion are also good at being funny. Some research corroborates this assumption: Howrigan and MacDonald (2008) and Köhler and Ruch (1996) both found small significant correlations between extraversion and humor production ($r_s = 0.17$ to 0.19), suggesting that funny people are more extraverted. But other research finds either a marginally significant relationship in the opposite direction (i.e., Greengross et al., 2011; Moran et al., 2014), or, as in the data presented here, fails to find a significant relationship at all (i.e. Feingold & Mazzella, 1993; Koppel & Sechrest, 1970). While this inconsistency could be attributed to noisy data, it seems more likely that extraversion has a nuanced relationship with humor production. For example, while open people tend to be better at writing jokes and captions, extraverted people probably shine in the delivery of humorous material to an audience. When a literature shows inconsistent findings for effects that are relatively small, meta-analysis is an illuminating tool. Low power and sampling variability are probably obscuring interesting effects in the literature as a whole, so a meta-analysis could reveal significant but small effects of traits other than openness to experience.

Practical Implications for Humor Assessment

One clear practical implication is that researchers should feel comfortable about expanding beyond the classic cartoon captions task. That task has worked well over the years, but it is wise to avoid putting all of one's empirical eggs in a single psychometric basket. Apart from diversifying, new tasks can capture shades of humor ability not captured by the cartoon task, which is relatively more visual, and possibly reduce the amount of time spent administering and scoring. Obviously, future work will need to

examine the newer family of tasks further—and develop even more new tasks—but these tasks appear to work well.

One limitation of the new tasks is that they are all essentially verbal. They assess wit conveyed via written language. Obviously, written language is a powerful medium for humor in everyday life, but a broader portfolio of humor tasks would serve researchers well. As the divergent thinking literature shows, verbal and visual tasks seem to tap different sets of skills and strategies (Clapham, 2004; Jankowska & Karwowski, 2015), so it's worth exploring non-verbal humor tasks. We haven't explored non-verbal tasks yet ourselves, and it is a great opportunity for researchers interested in expanding the science of humor research.

We should emphasize that we see our new tasks as supplementing rather than supplanting the long-standing captions task. Researchers should use several different tasks and pool them, such as with latent variable models, instead of trying to sift through the options to isolate the “one best task.” Our results illustrate the value of using many tasks in a latent variable approach. In Study 1, for example, the simple correlations between openness to experience with the 3 humor tasks ranged from $r = .19$ to $r = .37$. But having three tasks affords estimating a latent humor variable that separates shared variance from task-specific variance, and the relationship between openness to experience and the latent humor variable was much higher, $\beta = .48$. The virtue of developing new tasks, in our view, is to expand our portfolio of assessment tools.

The humors ratings were low overall. For all the items, the ratings clustered near the floor—the 1s and 2s mightily outnumbered the 4s and 5s. A virtue of a many-facet Rasch approach is that it estimates and corrects for the “severity” of the raters. Some raters were more severe than others—it doesn't take as much for some raters to find something funny—but the Rasch scaling approach suggests that the low scores are not merely a case of stern, implacable raters with bad senses of humor. Instead, most responses just weren't very funny. These tasks are fairly hard to do—being funny on the spot isn't

easy for most people—and they raise many of the same issues that divergent thinking tasks raise. For example, interference from obvious, salient ideas—such as responding to the “boring class” joke item with clichés like “watching grass grow” and “watching paint dry”—makes it hard to generate clever responses for most people. It would be illuminating to unpack the strategies that people use when they tackle these tasks, along the lines of Gilhooly, Fioratou, Anthony, and Wynn’s (2007) analysis of the strategies people use for unusual uses tasks.

The present studies used between 2 to 5 raters. All else equal, more raters should be better (Eckes, 2011), but constraints in time, personnel, and cost require some judgment about how many raters are necessary. Past humor research has used an unusually wide range of raters: from 2 (Feingold & Mazzella, 1991, 1993; Masten, 1986) to 6 (Greengross & Miller, 2011) to 12 (Köhler & Ruch, 1996; Kozbelt & Nishioka, 2010), 13 (Babad, 1974), and even 20 raters (Koppel & Sechrest, 1970). Over the years the three of us have scored an unholy number of open-ended responses for creativity, primarily responses to divergent thinking tasks (e.g., Beaty, Silvia, Nusbaum, Jauk, & Benedek, 2014; Beaty, Silvia, Nusbaum, & Vartanian, 2013; Nusbaum & Silvia, 2011; Silvia, Beaty, & Nusbaum, 2013; Silvia, Beaty, Nusbaum, Eddington, & Kwapil, 2014; Silvia, Martin, & Nusbaum, 2009; Silvia, Nusbaum, & Beaty, 2015; Silvia, Nusbaum, Berg, Martin, & O’Connor, 2009) and metaphor tasks (e.g., Beaty & Silvia, 2013; Silvia & Beaty, 2012). Our subjective impression of these humor production tasks is that they have relatively more variability due to raters than divergent thinking tasks do. Raters’ scores will vary more from each other because of different senses of humor and different thresholds for finding something funny. As a result, humor production tasks probably need more raters than divergent thinking tasks do. Our subjective impressions are consistent with the pattern of Rasch reliability estimates, which were higher for the studies with more raters.

Adding raters raises issues of efficiency, especially when a large sample of participants is given many open-ended items. As Primi (2014) points out, a virtue of applying many-facet Rasch analyses is

that the rating data can be incomplete: every rater need not rate every response. Such approaches are common in large-scale assessments, such as high-stakes language testing (Eckes, 2011), but are rarely seen in creativity research, which typically has every rater rate every response for every participant. Nevertheless, incomplete rating designs could be a fruitful way for researchers to increase the number of raters while reducing the typical burden on each one, and they deserve attention in future work.

References

- Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*(4), 340-345.
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., & ... De Raad, B. (2004a). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356-366.
- Babad, E. Y. (1974). A multi-method approach to the assessment of humor. *Journal of Personality, 42*, 618-631.
- Barbot, B., Tan, M., Randi, J., Santo-Donato, G., & Grigorenko, E. L. (2012). Essential skills for creative writing: Integrating multiple domain-specific perspectives. *Thinking Skills and Creativity, 7*, 209-223.
- Beatty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory and Cognition, 41*(2), 255-267. doi:10.3758/s13421-012-0258-5
- Beatty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory & Cognition, 42*, 1186-1197.
- Beatty, R. E., Silvia, P. J., Nusbaum, E. C., & Vartanian, O. (2013). Tired minds, tired ideas? Exploring insomnia and creativity. *Thinking Skills And Creativity, 9*, 69-75. doi:10.1016/j.tsc.2013.03.004
- Boyle, G. J., & Joss-Reid, J. M. (2004). Relationship of humor to health: A psychometric investigation. *British Journal of Health Psychology, 9*(1), 51-66. doi:10.1348/135910704322778722
- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag.
- Bressler, E. R., & Balshine, S. (2006). The influence of humor on desirability. *Evolution and Human Behavior, 27*(1), 29-39. doi:10.1016/j.evolhumbehav.2005.06.002
- Bressler, E. R., Martin, R. A., & Balshine, S. (2006). Production and appreciation of humor as sexually selected traits. *Evolution and Human Behavior, 27*(2), 121-130.

doi:10.1016/j.evolhumbehav.2005.09.001

Casteel, S. (2012). *Underwater dogs*. Boston, MA: Little, Brown, & Company.

Celso, B. G., Ebener, D., & Burkhead, E. (2003). Humor coping, health status, and life satisfaction among older adults residing in assisted living facilities. *Aging & Mental Health, 7*(6), 438-445.

doi:10.1080/13607860310001594691

Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and creative interest inventories. *Educational and Psychological Measurement, 64*, 828-841.

Conner, T. S., & Silvia, P. J. (2015). Creative days: A daily diary study of emotion, personality, and everyday creativity. *Psychology of Aesthetics, Creativity, and the Arts, 9*, 463-470.

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science, 24*, 299-349.

Cronbach, L.J., Nageswari, R., & Gleser, G.C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology, 16*, 137-163.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Dagnes, A. (2012). *A conservative walks into a bar: The politics of political humor*. New York, NY: Palgrave Macmillan.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of openness/intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality, 73*(4), 825-858. doi:10.1111/j.1467-6494.2005.00330.x

DeYoung, C. G., Quilty, L. C., Peterson, J. B., & Gray, J. R. (2014). Openness to experience, intellect, and cognitive ability. *Journal of Personality Assessment, 96*(1), 46-52.

doi:10.1080/00223891.2013.806327

Dyck, K. H., & Holtzman, S. (2013). Understanding humor styles and well-being: The importance of social

- relationships and gender. *Personality and Individual Differences*, 55(1), 53-58.
doi:10.1016/j.paid.2013.01.023
- Earleywine, M. (2010). *Humor 101*. New York, NY: Springer Publishing.
- Eckes, T. (2011). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments*. Frankfurt am Main, Germany: Peter Lang.
- Feingold, A. (1983). Measuring humor ability: Revision and construct validation of the Humor Perceptiveness Test. *Perceptual and Motor Skills*, 56(1), 159-166.
doi:10.2466/pms.1983.56.1.159
- Feingold, A., & Mazzella, R. (1991). Psychometric intelligence and verbal humor ability. *Personality and Individual Differences*, 12, 427-435.
- Feingold, A., & Mazzella, R. (1993). Preliminary validation of a multidimensional model of wittiness. *Journal of Personality*, 61(3), 439-456. doi:10.1111/j.1467-6494.1993.tb00288.x
- Fry, W. F. (1994). The biology of humor. *Humor: International Journal of Humor Research*, 7, 111-125.
- Fürst, G., Ghisletta, P., & Lubart, T. (in press). Toward an integrative model of creativity and personality: Theoretical suggestions and preliminary empirical testing. *Journal of Creative Behavior*.
doi: 10.1002/jocb.71
- Gilhooly, K. J., Fioratou, E. E., Anthony, S. H., & Wynn, V. V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, 98, 611-625.
- Greengross, G., Martin, R. A., & Miller, G. (2011). Personality traits, intelligence, humor styles, and humor production ability of professional stand-up comedians compared to college students. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 74-82. doi:10.1037/a0025774
- Greengross, G., & Miller, G. (2011). Humor ability reveals intelligence, predicts mating success, and is higher in males. *Intelligence*, 39, 188-192. doi:10.1016/j.intell.2011.03.006

- Howrigan, D. P., & MacDonald, K. B. (2008). Humor as a mental fitness indicator. *Evolutionary Psychology, 6*(4), 625-666.
- Jankowska, D. M., & Karwowski, M. (2015). Measuring creative imagery abilities. *Frontiers in Psychology, 6* (1591).
- Joy, S. (2004). Innovation motivation: The need to be different. *Creativity Research Journal, 16*, 313-330.
- Karwowski, M., & Lebuda, I. (2016). The big five, the huge two, and creative self-beliefs: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 10*, 214-232.
- Kaufman, S. B. (2013). Opening up openness to experience: A four-factor model and relations to creative achievement in the arts and sciences. *Journal of Creative Behavior, 47*, 233-255.
- Kaufman, S. B., Kozbelt, A., Bromley, M. L., & Miller, G. R. (2008). The role of creativity and humor in human mate selection. In G. Geher, G. Miller, G. Geher, G. Miller (Eds.) , *Mating intelligence: Sex, relationships, and the mind's reproductive system* (pp. 227-262). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Köhler, G., & Ruch, W. (1996). Sources of variance in current sense of humor inventories: How much substance, how much method variance?. *Humor: International Journal of Humor Research, 9*, 363-397. doi:10.1515/humr.1996.9.3-4.363
- Koppel, M. A., & Sechrest, L. (1970). A multitrait-multimethod matrix analysis of sense of humor. *Educational and Psychological Measurement, 30*, 77-85. doi:10.1177/001316447003000107
- Kozbelt, A., & Nishioka, K. (2010). Humor comprehension, humor production, and insight: An exploratory study. *Humor: International Journal of Humor Research, 23*(3), 375-401. doi:10.1515/HUMR.2010.017
- Kuiper, N. A., & Nicholl, S. (2004). Thoughts of feeling better? Sense of humor and physical health. *Humor: International Journal of Humor Research, 17*, 37-66. doi:10.1515/humr.2004.007
- Lee, K. & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory.

- Multivariate Behavioral Research*, 39, 329-358.
- Lee, K., & Ashton, M. C. (2014). The dark triad, the big five, and the HEXACO model. *Personality and Individual Differences*, doi:10.1016/j.paid.2014.01.048
- Linacre, J. M. (1997). KR-20/Cronbach Alpha or Rasch person reliability: Which tells the "truth"? *Rasch Measurement Transactions*, 11, 580-581.
- Linacre, J. M. (2014). *FACETS 3.71.4* [Computer software]. Chicago, IL: Winsteps.
- Martin, R. A. (2004). Sense of humor and physical health: Theoretical issues, recent findings, and future directions. *Humor: International Journal of Humor Research*, 17, 1-19.
doi:10.1515/humr.2004.005
- Masten, A. S. (1986). Humor and competence in school-aged children. *Child Development*, 57(2), 461-473. doi:10.2307/1130601
- McCrae, R. R., & Costa Jr, P. T. (2007). Brief versions of the NEO-PI-3. *Journal of Individual Differences*, 28(3), 116.
- McCrae, R. R., & Costa, P. J. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, L. A. Pervin, O. P. John, R. W. Robins, L. A. Pervin (Eds.), *Handbook of personality: Theory and research (3rd ed.)* (pp. 159-181). New York, NY, US: Guilford Press.
- Moran, J. M., Rain, M., Page-Gould, E., & Mar, R. A. (2014). Do I amuse you? Asymmetric predictors for humor appreciation and humor production. *Journal of Research in Personality*, 49, 8-13.
doi:10.1016/j.jrp.2013.12.002
- Nusbaum, E. C. & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, 39, 36-45.
- Nusbaum, E. C., & Silvia, P. J. (in press). What are funny people like? Exploring the crossroads of humor ability and openness to experience. In G. J. Feist, R. Reiter-Palmon, & J. C. Kaufman (Eds.), *Cambridge handbook of creativity and personality research*. New York, NY: Cambridge University

Press.

- Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2014). Ready, set, create: What instructing people to “be creative” reveals about the meaning and mechanisms of divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 423-432. doi:10.1037/a0036549
- Nusbaum, E. C., & Silvia, P. J. (2014). Unusual aesthetic states. In P. P. L. Tinio & J. K. Smith (Eds.), *The Cambridge handbook of the psychology of aesthetics and the arts* (pp. 519-539). Cambridge, UK: Cambridge University Press.
- Nusbaum, E. C., Silvia, P. J., Beaty, R. E., Burgin, C. J., Hodges, D. A., & Kwapil, T. R. (2014). Listening between the notes: Aesthetic chills in everyday music listening. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 104-109.
- Oleynick, V. C., DeYoung, C. G., Hyde, E., Kaufman, S. B., Beaty, R. E., & Silvia, P. J. (in press). Openness/Intellect: The core of the creative personality. In G. J. Feist, R. Reiter-Palmon, & J. C. Kaufman (Eds.), *Cambridge handbook of creativity and personality research*. New York, NY: Cambridge University Press.
- Polimeni, J. O., Campbell, D. W., Gill, D., Sawatzky, B. L., & Reiss, J. P. (2010). Diminished humour perception in schizophrenia: Relationship to social and cognitive functioning. *Journal of Psychiatric Research, 44*(7), 434-440. doi:10.1016/j.jpsychires.2009.10.003
- Primi, R. (2014). Divergent productions of metaphors: Combining many-facet Rasch measurement and cognitive psychology in the assessment of creativity. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 461-474.
- Silvia, P. J. (2007). Knowledge-based assessment of expertise in the arts: Exploring aesthetic fluency. *Psychology of Aesthetics, Creativity, and the Arts, 1*(4), 247-249. doi:10.1037/1931-3896.1.4.247
- Silvia, P. J., & Beaty, R. E. (2012). Making creative metaphors: The importance of fluid intelligence for creative thought. *Intelligence, 40*(4), 343-351. doi:10.1016/j.intell.2012.02.005

- Silvia, P. J., Beaty, R. E., & Nusbaum, E. C. (2013). Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence, 41*(5), 328-340. doi:10.1016/j.intell.2013.05.004
- Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., Levin-Aspenson, H., & Kwapil, T. R. (2014). Everyday creativity in daily life: An experience-sampling study of “little c” creativity. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 183-188.
- Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., & Kwapil, T. R. (2014). Creative motivation: Creative achievement predicts cardiac autonomic markers of effort during divergent thinking. *Biological Psychology, 102*, 30-37. doi:10.1016/j.biopsycho.2014.07.010
- Silvia, P. J., Martin, C., & Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity, 4*(2), 79-85. doi:10.1016/j.tsc.2009.06.005
- Silvia, P. J., Nusbaum, E. C., & Beaty, R. E. (2015). Old or new? Evaluating the old/new scoring method for divergent thinking tasks. *Journal of Creative Behavior*, doi:10.1002/jocb.101
- Silvia, P. J., Nusbaum, E. C., Berg, C., Martin, C., & O'Connor, A. (2009). Openness to experience, plasticity, and creativity: Exploring lower-order, high-order, and interactive effects. *Journal Of Research In Personality, 43*(6), 1087-1090. doi:10.1016/j.jrp.2009.04.015
- Sprecher, S., & Regan, P. C. (2002). Liking some things (in some people) more than others: Partner preferences in romantic relationships and friendships. *Journal of Social and Personal Relationships, 19*(4), 463-481. doi:10.1177/0265407502019004048
- Stanley, J. T., Lohani, M., & Isaacowitz, D. M. (2014). Age-related differences in judgments of inappropriate behavior are related to humor style preferences. *Psychology and Aging, 29*(3), 528-541. doi:10.1037/a0036666
- Svebak, S. (2005). Does humor benefit health and well-being? *Journal of the Norwegian Psychological*

- Association*, 42(10), 909-912.
- Swami, V., & Furnham, A. (2014). Personality and aesthetic experiences. In P. P. L. Tinio & J. K. Smith (Eds.), *The Cambridge handbook of the psychology of aesthetics and the arts* (pp. 540-561). Cambridge, UK: Cambridge University Press.
- Terracciano, A., McCrae, R. R., Hagemann, D., & Costa, P. T., Jr. (2003). Individual difference variables, affective differentiation, and the structures of affect. *Journal of Personality*, 71, 669-703.
- Turner, R. G. (1980). Self-monitoring and humor production. *Journal of Personality*, 48(2), 163-172. doi:10.1111/j.1467-6494.1980.tb00825.x
- Watson, D. (2000). *Mood and temperament*. New York: Guilford.
- Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin*, 50, 362-370. doi:10.1037/h0060857
- Wilt, J., & Revelle, W. (2009). Extraversion. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 27-45). New York, NY: Guilford.
- Wisniewski, E. J. (1997). Conceptual combination: Possibilities and esthetics. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 51-81). Washington, DC: American Psychological Association.
- Ziegler, M., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality*, 46(2), 173-183. doi:10.1016/j.jrp.2012.01.002
- Zingrone, N. L., Alvarado, C. S., & Dalton, K. (1998). Psi experiences and the "Big Five": Relating the NEO-PI-R to the experience claims of experimental subjects. *European Journal of Parapsychology*, 14, 31-51.

Table 1

Descriptive Statistics: Study 1

Variable	<i>M</i>	<i>Variance</i>	1	2	3	4	5	6	7	8
1. Résumés	1.49	.16	1							
2. Jokes	1.55	.12	.41	1						
3. Captions	1.40	.11	.37	.44	1					
4. O	3.47	.25	.30	.37	.19	1				
5. C	3.63	.30	-.14	-.07	-.03	.04	1			
6. E	3.61	.28	.04	.02	-.04	.17	.29	1		
7. A	3.60	.27	.07	.04	.04	.16	.21	.10	1	
8. N	3.08	.46	-.03	-.02	-.12	.00	-.31	-.41	-.04	1

Note. $n = 166$. The scores for Captions, Jokes, and Résumés are “fair average” scores estimated from many-facet Rasch models.

Table 2

Effects of Personality on Humor Ability: Summary of Studies 1, 2, and 3

	Study 1		Study 2		Study 3	
	β	95% CI	β	95% CI	β	95% CI
Openness to Experience	.48	.34, .63	.54	.39, .70	.36	.22, .50
Conscientiousness	-.18	-.37, .01	-.09	-.28, .09	-.11	-.27, .04
Extraversion	-.09	-.30, .13	-.04	-.22, .15	-.04	-.18, .11
Agreeableness	.04	-.14, .22	.15	-.04, .34	.14	-.03, .30
Neuroticism	-.17	-.35, .02	.04	-.14, .23	-.05	-.19, .09
Honesty-Humility					-.07	-.23, .09

Note. The effects are standardized regression weights. In each model, the predictor variables covaried freely with each other. In Study 3, Neuroticism refers to the HEXACO Emotionality Factor, and Extraversion refers to the HEXACO eXtraversion factor.

Table 3

Descriptive Statistics: Study 2

Variable	<i>M</i>	<i>Variance</i>	1	2	3	4	5	6	7
1. Jokes	1.52	.16	1						
2. Captions	1.49	.17	.42	1					
3. O	3.71	.29	.27	.46	1				
4. C	3.50	.28	-.15	-.20	-.29	1			
5. E	3.63	.24	-.07	.02	.00	.14	1		
6. A	3.64	.26	.12	.06	-.01	.00	.18	1	
7. N	3.09	.34	.18	.10	.21	-.38	-.25	.01	1

Note. $n = 147$. The scores for Jokes and Captions are “fair average” scores estimated from many-facet Rasch models.

Table 4

Descriptive Statistics: Study 3

Variable	<i>M</i>	<i>Variance</i>	1	2	3	4	5	6	7
1. Humor Ability	1.51	.15	1						
2. H	3.18	.30	.01	1					
3. E	3.47	.25	-.06	-.02	1				
4. X	3.45	.37	-.04	.14	-.09	1			
5. A	2.97	.27	.17	.30	-.06	.08	1		
6. C	3.45	.30	-.10	.24	.04	.38	.20	1	
7. O	3.12	.28	.36	.20	-.01	.10	.21	.06	1

Note. $n = 138$. In the HEXACO model, X is Extraversion and E is Emotionality (Neuroticism). The humor ability score is a “fair average” score estimated from a many-facet Rasch model for 2 raters, 3 tasks, and 3 items per task.

Figure 1. Effects of the Big Five factors on humor ability: Study 1

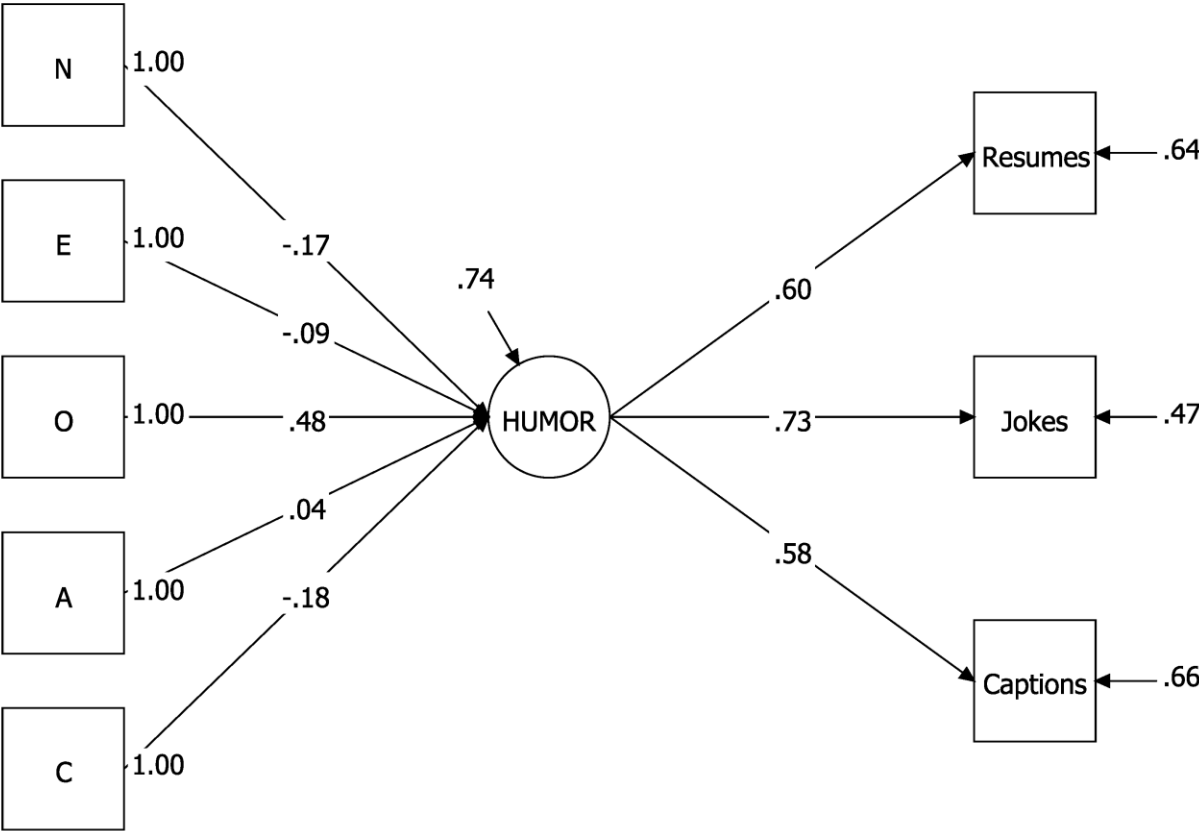


Figure 2. Effects of the Big Five factors on humor ability: Study 2

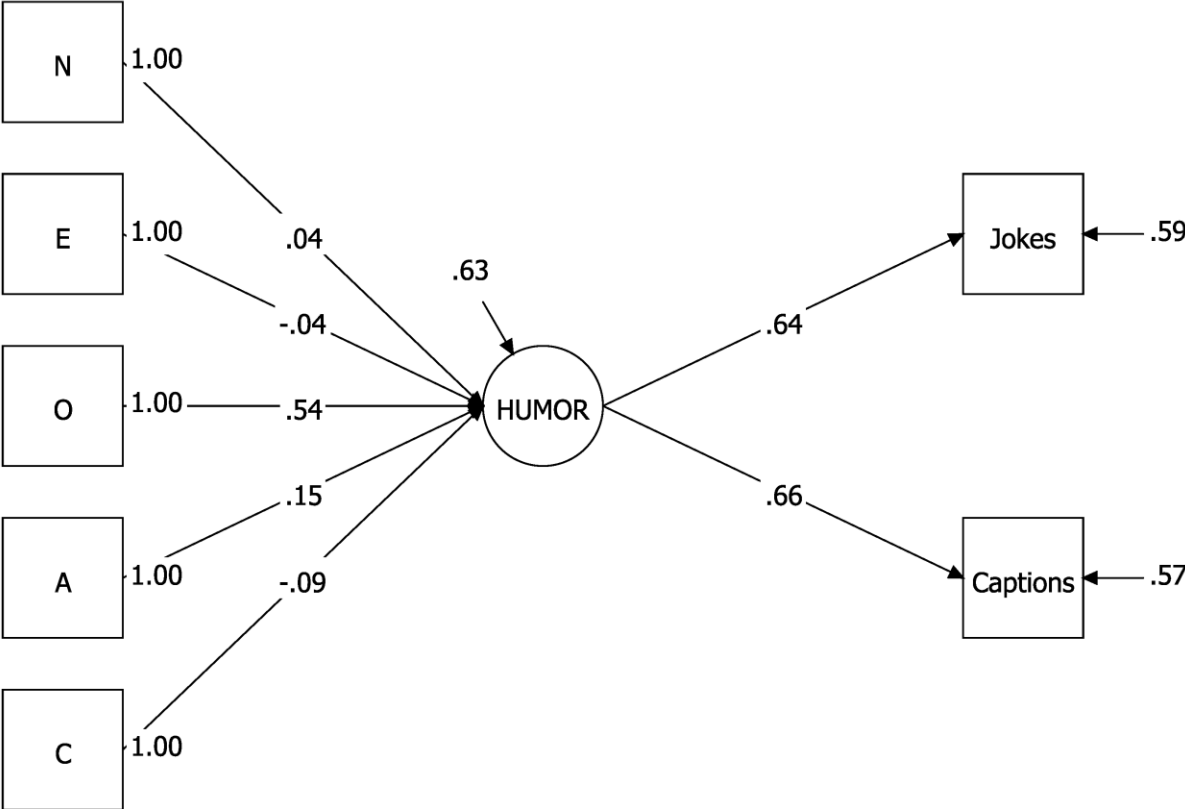
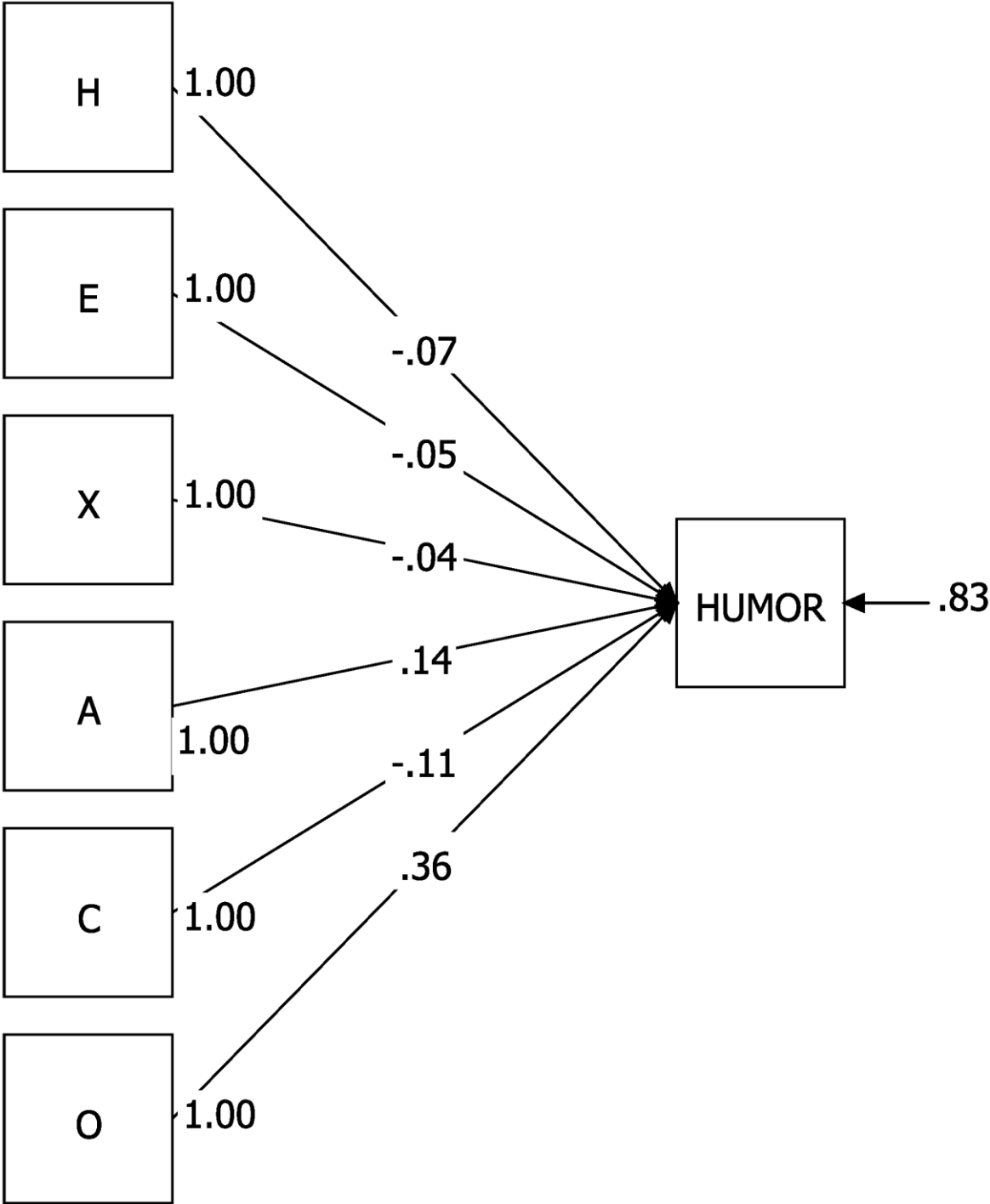


Figure 3. Effects of the HEXACO factors on humor ability: Study 3



Appendix: Joke Stems

1. *Imagine that one of your friends wants your opinion on how well she sings. She sings a minute or two to demonstrate her voice, and you cringe—she might be the worst singer you’ve ever heard. When she asks, “So how was it?” you decide to be honest, so you say, “To be honest, listening to that was like...”*

Please complete the phrase “To be honest, listening to that was like...” with something funny below.

2. *Imagine that one of your classes this semester is incredibly boring, and you’re trying to convey just how boring this class is to one of your friends. So you say, “Seriously, that class is so boring...”* Please

complete the phrase “Seriously, that class is so boring...” with something funny below.

3. *Imagine that your friend invites you over and cooks dinner—and the food is totally horrible and disgusting. Later, when describing it to someone else, you say, “Wow, that food was so bad...”* Please

complete the phrase “Wow, that food was so bad...” with something funny below.