



HAL
open science

Getting students interested in psychological measurement by experiencing the Barnum effect

Corentin Gonthier, Noémie Thomassin

► **To cite this version:**

Corentin Gonthier, Noémie Thomassin. Getting students interested in psychological measurement by experiencing the Barnum effect. *Teaching of Psychology*, 2024, 10.1177/00986283241240454 . hal-04685596

HAL Id: hal-04685596

<https://hal.science/hal-04685596v1>

Submitted on 3 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Getting Students Interested in Psychological Measurement
by Experiencing the Barnum Effect**

Corentin Gonthier^{1,2}, Noémie Thomassin

¹ Nantes Université, LPPL UR 4638, Nantes, France

² Institut Universitaire de France

Correspondence concerning this article should be addressed to Corentin Gonthier,
Laboratoire LPPL, Chemin de la Censive du Tertre, BP 81227, 44312 Nantes Cedex 3,
France

E-mail: corentin.gonthier@univ-nantes.fr

Abstract

Background: Fake psychological tests tend to be viewed as completely believable (Barnum effect), meaning psychometric properties cannot be judged subjectively. Experiencing this effect first-hand could help get students interested in the science of psychological assessment.

Objective: In a blended learning perspective, we created materials to elicit a Barnum effect with a brief online fake test and tested the benefits of this procedure for student engagement.

Method: Five cohorts of undergraduate psychology students (total N = 2,269) completed the fake test online prior to their first psychometrics course. We collected single-item ratings of perceived effectiveness for the test, willingness to use it, and course engagement (interest and enthusiasm for the course). We also retrieved grades.

Results: Eighty-five percent of students rated the effectiveness of the fake test above the neutral point; seventy-seven percent were willing to use it as future psychologists.

Experiencing the effect was related to significantly increased interest and enthusiasm, and higher grades.

Conclusion: Having students experience the Barnum effect for themselves is a useful and effortless approach to get them interested in psychometrics.

Teaching implications: Teachers of introductory courses related to psychological measurement should consider having all students experience the Barnum effect. All materials are provided.

Keywords: Psychometrics; Forer effect; Barnum effect; Student interest; Teaching experiment

Getting Students Interested in Psychological Measurement by Experiencing the Barnum Effect

Psychometrics, or the science of psychological measurement, is an essential part of the curriculum of studies in psychology (Chappell et al., 2021) and a required topic in accredited programs for both the American Psychological Association (APA, 2023) and British Psychological Society (2019). When teaching psychometrics, it is sometimes difficult to find material to engage students, as the field includes many dry topics (from the interplay between reliability and validity as a function of ability level to exploratory factor analysis, through item response theory). This is true even for courses on individual differences which touch on psychological measurement, where teachers must convince students that measurement matters and that not just any test can be used to measure constructs such as personality or intelligence. The Barnum effect, or Forer effect, appears to be a promising way to convey the importance of this topic because it is something concrete, that students can experience for themselves.

The Barnum effect was first described by Forer (1949, 1968; Meehl, 1956) and can be summarized as follows: people tend to believe in the results of psychological tests. When receiving feedback from a (personality) test, subjects tend to find that the profile they are given describes them well – even when this feedback is completely bogus. The Barnum effect can be viewed as a foundation of psychometrics: tests must be developed, and their psychometric qualities assessed, through a rigorous procedure, because it is impossible to tell at first glance whether a test functions well: test users will always tend to be satisfied in the test whether it works or not.

In his classic work, Bertram Forer (1949) had a sample of 39 students complete a series of 279 items enquiring about their hobbies, hopes, ambitions, etc. He pretended to rate the results of this test and then delivered bogus feedback to students. This feedback was

mainly based on horoscopes taken in an astrology book, chosen for their nearly universal applicability; critically, the feedback was identical for all students, and thus completely non-informative. The results showed that students overwhelmingly believed that the test had been effective in revealing their personality (38 out of 39 students scored the adequacy of the test at 4 or 5 on a scale from 0 to 5).

Since this initial demonstration, the Barnum effect has been replicated numerous times (for reviews, see Dickson & Kelly, 1985; Furnham & Schofield, 1987). Studies have rarely failed to find a Barnum effect (Furnham & Schofield, 1987). Not only do participants tend to believe in bogus personality profiles, but they are also often unable to distinguish between bogus feedback and their actual profile on a test (Dies, 1972; Merrens & Richards, 1970; Sundberg, 1955), and their trust in the effectiveness of personality tests even tends to increase after receiving a bogus profile (Hinrichsen & Bradley, 1974). The effect seems to work well regardless of gender or personality (Furnham & Schofield, 1987), despite possible moderation by some personality traits (e.g., Poškus, 2014).

The fact that the Barnum effect functions so well and is so stable across studies and across individuals makes it highly suitable as a teaching tool, as the effect is expected to appear for most students. Furthermore, the fact that the effect quickly and directly challenges the participants' assumptions (Forer, 1949), like revealing an optical illusion, makes it ideal to introduce psychometrics by provoking a cognitive conflict. Such a cognitive conflict should serve to actively involve students and change their preconceptions about the interest of psychometrics (see Dreyfus et al., 1990; Hewson & Hewson, 1984; Nussbaum & Novick, 1982; Posner et al., 1982). In fact, the procedure of presenting bogus personality profiles to students to directly confront them with their inability to judge subjectively the qualities of a psychological test was explicitly designed to teach them importance of psychometric validation (Forer, 1949). We have repeatedly done this in our own courses; like Forer (1949),

displaying a slide with the bogus personality profile and asking students "do you recognize this personality profile?" elicited bursts of laughter in the class, and students anecdotally reported being impressed that the effect had worked so well on them.

Since Forer's study, a few authors have reported using the effect for teaching (e.g., Bachrach & Pattishall, 1960; Beins, 1993; Boyce & Geller, 2002), but there seem to be two major barriers to more widespread adoption of the Barnum effect as a teaching device. The first is the lack of an updated procedure, with easily accessible materials, suitable for large-scale teaching. For instance, Forer (1949) used a 279-items-long test and wrote the name of each student on their printed personality profile. This procedure poses a problem of acceptability for both students (i.e., has a long completion time that is liable to discourage students, especially for a course that is perceived as unattractive in the first place) and teachers (i.e., Forer's class included 39 students; printing and handwriting student names on each report becomes impractical and time-consuming for larger class sizes). The second barrier to adoption is that the actual impact of experiencing the effect on student engagement has not been tested. At least four major empirical questions remain open.

The first open empirical question is whether the Barnum effect can still work in contemporary students, who may be more skeptical than students of the 20th century due to the omnipresence of bogus personality tests available online (it is noteworthy that almost all studies on the Barnum effect date from before 1985). While we are not aware of any research specifically regarding trust in psychological tests, cohort studies have also shown a trend towards declining trust (including trust in science, which could affect tests presented as "scientifically validated") and increasing cynicism in recent years (e.g. Gauchat, 2012; Trzesniewski & Donnellan, 2010; Twenge et al., 2012). Relatedly, it is unknown whether the effect can work over consecutive years of teaching: contemporary students have easy means of communicating online, potentially informing others about the deception. We have seen, on

several occasions, students refusing to engage with the fake test, because they had been warned of the deception by more senior students: if effect size decreased across teaching years due to students communicating about the deception, this would make the procedure entirely useless as a teaching device.

The second question is whether students perceive the Barnum effect as relevant to their future profession as test users: later research has shown that psychology students, despite rating a bogus personality profile as applying to them very well, can detect that the profile is trivial (Greene, 1977). This means they may not consider the fake test as effective, potentially downplaying the significance of the effect as future professionals. The third question is whether experiencing the Barnum effect directly actually does increase student engagement in the course and their willingness to engage with the material (see Fredricks et al., 2004); being deceived could potentially have the opposite effect. The fourth question is whether experiencing the Barnum effect is in fact related to higher student grades, which could potentially reflect a beneficial effect on student achievement.

Current Study

The overarching goal of the current study was to promote adoption of the procedure imagined by Forer (1949) to increase student willingness to study psychological measurement. This encompassed two objectives. The first was to provide suitable teaching materials, including an updated fake test and personality profile that could be used with little effort from the teacher. The second was to complement prior literature (mainly focused on the existence of the Barnum effect and its boundary conditions) by testing the possibility of eliciting a Barnum effect across consecutive years of teaching, in a way that students identified as relevant to their training; eliciting increased student engagement in the course; and ultimately promoting student achievement.

The first objective involved creating a new version of the materials devised by Forer (1949). Given that the Barnum effect appears to work identically for various types of tests (Hinrichsen & Bradley, 1974; Weinberger & Bradley, 1980), whether the test is presented as scored by a computer or a human (Baillargeon & Danis, 1984; Orpen & Jamotte, 1975; Snyder & Larson, 1972), whether feedback is oral or written (Snyder & Shenkel, 1976), and whether the test is short or long (Merrens & Richards, 1973), the effect seemed perfectly suitable to a much shorter fake test intended to be completed online – in the spirit of blended learning. Forer's bogus personality profile already met the constraints described in later research and was only slightly updated. We also wrote step-by-step advice for implementation, which can be accessed online (Gonthier, 2024).

The second objective required empirically testing the new test materials and the effect of the procedure on student engagement. To this end, we used the materials with students who completed the test prior to their first psychometrics course, over 5 years of teaching. We sought to answer the four questions outlined above: (1) Do the new test materials succeed at eliciting a substantial Barnum effect in students in the context of actual teaching? This included three related questions: (1a) Do students rate the fake test as effective above the neutral point, despite the test being automated and much shorter than the original materials (Forer, 1949)? (1b) How do effectiveness ratings compare with those initially reported by Forer? (1c) Does the effect continue to appear across multiple consecutive years of teaching with these materials, despite the possibility of students communicating about the deception? (2) Does the experience appear relevant to the training of psychology students? In other words, do students rate their willingness to use the fake test in their capacity as future psychologists above the neutral point?

(3) Do students demonstrate more engagement in the course (e.g., interest in the course and enthusiasm about the course) after receiving their bogus personality profile and learning about the Barnum effect?

(4) Do students who have experienced the effect ultimately demonstrate higher grades?

Method

The procedure of this study was approved by the Nantes University institutional review board (agreement n°13102023). Complete course materials for a demonstration of the effect (including a step-by-step discussion of how to implement the fake test and how to discuss the effect in class) are available on Open Science Framework (Gonthier, 2024).

Participants

For 5 consecutive years, all 2nd-year students enrolled in a psychology major in a medium-sized French university were invited to complete the fake personality test online. None of the students had followed a course related to either personality or psychometrics before, although a few notions related to test design were introduced in their first year of study (e.g., how to word items in a questionnaire). Students were always invited to complete the test prior to their first psychometrics course, at the beginning of the second semester. The procedure was used for 2nd-year students before their first course because advanced psychology students may be a little less vulnerable to the Barnum effect (Greene, 1977; Greene et al., 1979).

Total sample size was $N = 2269$, including data collected for university years 2017-2018 ($n = 437$ respondents out of 646 students, 68% response rate), 2018-2019 ($n = 584$ respondents out of 665 students, 88% response rate), 2019-2020 ($n = 419$ out of 655 students, 64% response rate), 2020-2021 ($n = 511$ out of 778 students, 66% response rate), and 2021-2022 ($n = 318$ out of 568 students, 56% response rate). Because the test was fully anonymous, demographic data were not collected individually; the overall composition of the

student cohorts over these 5 years (as retrieved from the university database of psychology students) was 82% female, median age = 22 years.

Materials

Both the fake personality test and bogus personality profile are available on Open Science Framework (Gonthier, 2024), in English-speaking and French-speaking versions.

Fake Personality Test

The fake personality test was designed in keeping with the following constraints: it had to be computerized (with the possibility of being completed online), short (with an estimated completion time in the 5-10 minutes range to maximize student participation), and a little mysterious (with some studies suggesting that this can slightly increase the magnitude of the Barnum effect: Snyder, 1974; see also Furnham & Schofield, 1987). The test included two sections. The first comprised 30 items sampled from online personality tests, so that they resembled material familiar to students; for each item, respondents had to rate agreement on a 5-point Likert scale. The second section, which was expected to be less familiar to students and add some mystery, was taken from Forer (1948, 1949) and comprised 20 life goals; for each item, respondents had to report whether they recognized themselves in this life goal, with three response options (*not at all, somewhat, a lot*).

Bogus Personality Profile

The literature suggests that the bogus personality profile provided by the test should go in both directions (e.g., "you are generally cheerful but get depressed at times"), be favorable, be vague, and be usually true for the subject group (Sundberg, 1955; Dickson & Kelly, 1985). The profile historically used by Forer (1949) already followed these guidelines and was only slightly adapted for this study: we removed two statements (one regarding sexual adjustment, which could be troublesome for some students, and one regarding security as a life goal, which fit less well with the rest of the profile) and restructured the other

thirteen statements in five paragraphs loosely resembling the five main dimensions of personality, to increase credibility of the profile for students familiar with the Big Five model.

Measures

The following measures were collected to answer the four study questions:

- (1) Students from all five cohorts (2017-2018 to 2021-2022) rated the perceived effectiveness of the test, immediately after receiving the bogus profile, with the following prompt and response scale, identical to the princeps study to allow for direct comparison (Forer, 1949): "On a scale from 0 (*very bad*) to 5 (*perfect*), how effective was the test in revealing features of your personality?"
- (2) Students from cohorts 2019-2020, 2020-2021 and 2021-2022 additionally rated the perceived relevance of the test to their future profession as psychologists, with the following prompt: "On a scale from 0 (*not at all*) to 5 (*totally*), as a psychologist, would you consider using this test to assess the personality of a patient?"
- (3) Students from cohorts 2019-2020, 2020-2021 and 2021-2022 additionally provided before/after ratings of engagement in the course, using two items assessing interest and enthusiasm: "How interested are you in following a psychometrics course?" and "How glad are you that you have to follow a psychometrics course?". Both questions were rated on a scale from 1 (*not at all*) to 7 (*totally*). Both questions were asked immediately before taking the test, and again immediately after being debriefed about the Barnum effect.
- (4) Students from the 2020-2021 cohort reported their student IDs, allowing to cross-reference their responses with their grades on final exams at the end of the semester while preserving anonymity. Grades were retrieved for the psychometrics final exam, as well as for two other courses with similar modes of assessment (*Cognitive psychology* and *Experimental methodology and statistics*).

Procedure

Two weeks before the first session of their psychometrics course, students received an e-mail inviting them to complete the test online. The e-mail was sent by their teacher and the test was presented as "a validated, state-of-the art personality test" to reinforce credibility of the bogus personality profile (Dana & Graham, 1976; Halperin et al., 1976; Snyder & Shenkel, 1976). The invitation stressed that taking part in the test was not mandatory, and that responses would remain confidential. Students who chose to do so clicked a link to access the test on a university webpage.

Students completed the fake test, received the bogus personality profile, and answered the additional measures as detailed in the preceding section. Students were then immediately debriefed about the deception, using a short explanation of the Barnum effect presented on the following page on the website. This departs from the procedure used by Forer (1949) who only debriefed students in class; this change may conceivably decrease the impact of the demonstration, but it appeared necessary to debrief students as soon as possible (APA, 2017; Boynton et al., 2013) both to limit the extent of deception and to avoid the possibility of students never being debriefed if they missed the first session. Students were asked not to communicate about the deception with other students. The effect was discussed at length during the first session of the course. Details about the procedure and advice for course contents are available as online supplemental materials on Open Science Framework (Gonthier, 2024).

Data Analysis

Question 1 regarding the effectiveness of our implementation of the Barnum effect over multiple consecutive years was tested by comparing average perceived effectiveness ratings for the fake test to the neutral point of the rating scale (2.5 on a scale of 0 to 5) using a one-sample Student's *t*-test. The overall distribution of ratings was compared to Forer's

(1949) original results using the Kolmogorov distance. We also compared average ratings over years using a between-subjects ANOVA, complemented by a Bayes factor to test the null hypothesis of no difference across years.

Question 2 regarding the relevance of the Barnum effect to students' self-concept as future psychologists was tested by comparing average ratings of intention to use the test as psychologists to the neutral point of the rating scale, identical to Question 1. Of secondary interest, we also computed the bivariate correlation between perceived effectiveness of the fake test and intention to use it as a psychologist.

Question 3 regarding the impact of experiencing the Barnum effect on engagement in the course used within-subjects Student's *t*-tests (with Cohen's *d_z* as an effect size) to compare average ratings for interest and satisfaction before and after experiencing the Barnum effect. We also tested whether the increase of engagement varied as a function of the intensity of the Barnum effect (perceived effectiveness of the fake test) using an analysis based on the general linear model, with rating as the dependent variable, rating time (before vs. after) as a within-subjects categorical independent variable, and perceived effectiveness of the test as a between-subjects numerical independent variable, including the interaction between the two.

Lastly, Question 4 regarding the relation between experiencing the Barnum effect and performance on the final exam tested the difference in average grades between students who completed and students who did not complete the fake personality test using a between-subjects Student's *t*-test. To account for the possibility of sampling bias (which would occur if high-performing students with better grades overall were more likely to go through the effort of completing the fake personality test), we also conducted an analysis of covariance (ANCOVA) additionally controlling for grades on *Cognitive psychology* and *Experimental psychology and statistics*.

Results

Do the Test Materials Succeed at Eliciting a Stable Barnum Effect?

The results for the effectiveness rating of the fake personality test are reported in Figure 1. Overall, students appeared to be convinced by the fake test: a total of 85% of students rated the effectiveness of the test at 3 or higher (i.e., 4 or higher on a 6-point scale, since the rating scale used by Forer starts at 0) and 64% of students even rated the effectiveness at 4 or higher (i.e., 5 or higher on a 6-point scale). The average rating was $M = 3.59$ ($SD = 1.17$), significantly above the neutral point of 2.5, $t(2268) = 44.44$, $p < .001$, $d = 0.93$. In short, the Barnum effect appeared to function very well in this sample.

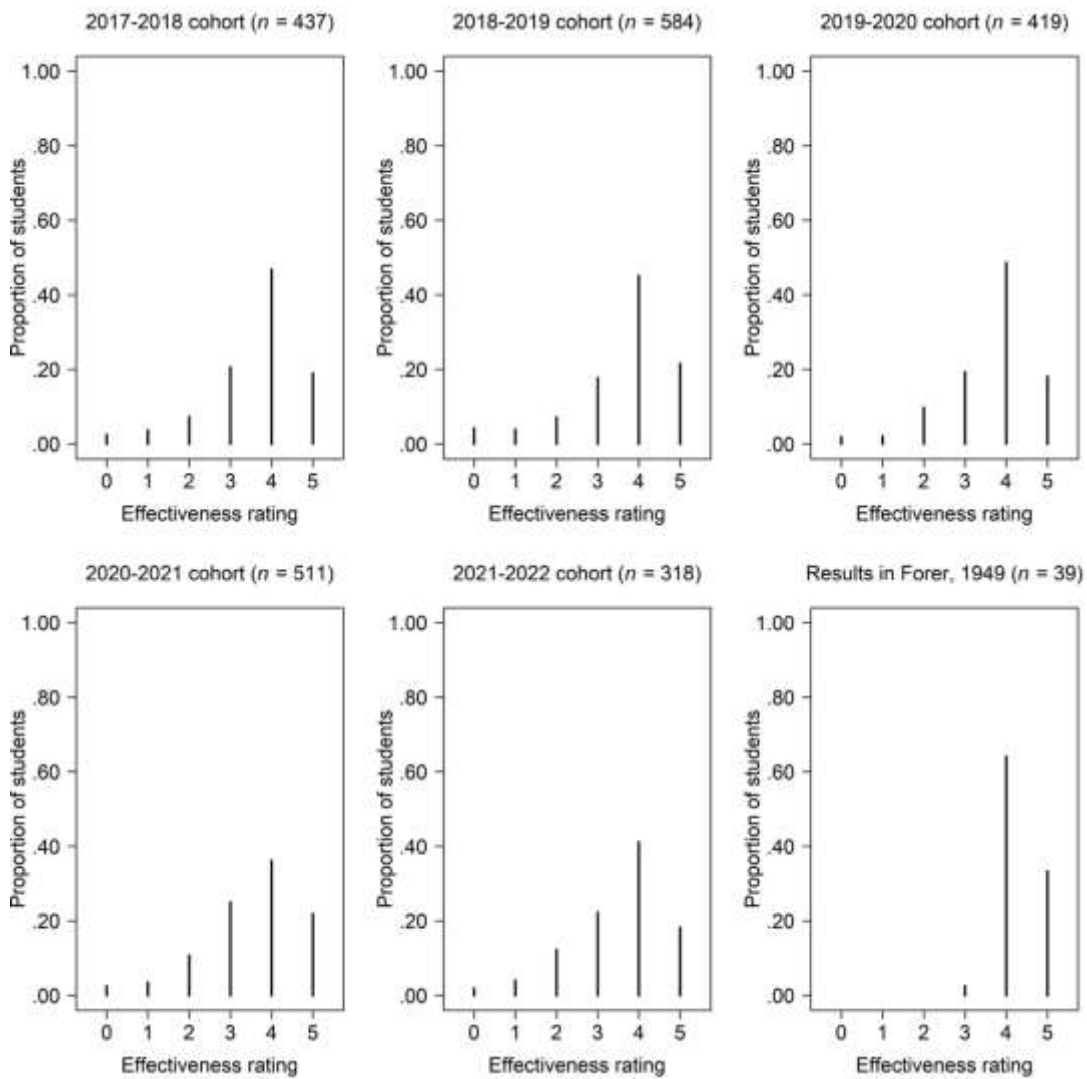
Ratings were a little less favorable in our study than they were in Forer (1949), where almost all students rated effectiveness at 4 or higher: there was a significant difference in the distribution of ratings between our sample and Forer's sample, $D = 0.34$, $p < .001$ (see Figure 1, bottom right quadrant). This could be due to participants being 2nd-year students instead of participants in an introductory psychology course, the use of a much shorter test, the online mode of completion, the test not being administered directly by the teacher, the feedback not being personalized with the written name of the student, inter-generational differences due e.g., to increased exposure to online tests, phrasing the question in a way slightly different from Forer (after translation, our phrasing was closer to "how effective was the test in revealing your personality" than to the original "how effective is the test in revealing personality"), or random chance in the small sample collected by Forer. However, with 85% of students rating the effectiveness of the fake test above the neutral point, these slightly lower ratings did not prevent the effect from working.

There was no difference between the five cohorts in the magnitude of the Barnum effect, $F(4, 2264) = 0.92$, $p = .451$, $\eta^2_p = .00$, $BF_{01} = 765.95$. This absence of difference confirms that using the Barnum effect as a teaching device across consecutive years did not

compromise the effect. In sum, the test materials succeeded at eliciting a substantial Barnum effect that was stable across multiple years, although a little weaker than in the original study by Forer (1949).

Figure 1

Effectiveness Rating for the Fake Personality Test as a Function of Cohort



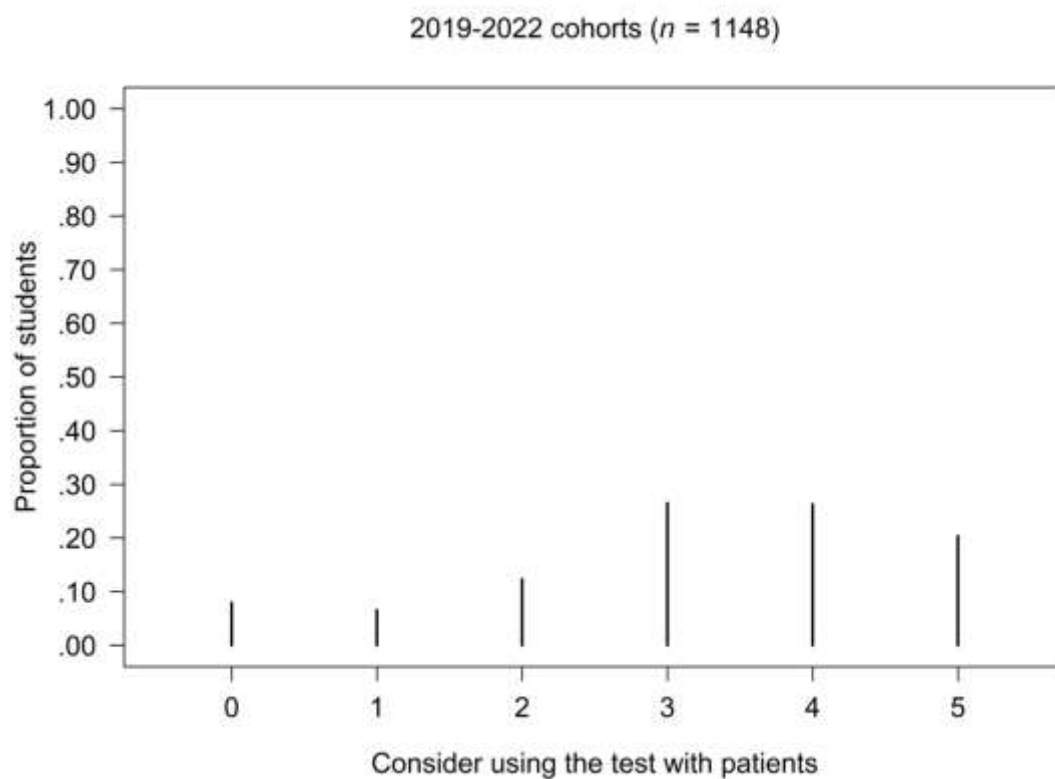
Note. All quadrants represent a cohort of students collected for this study, except the bottom-right quadrant, which displays the results reported by Forer (1949) for comparison.

Does the Experience Appear Relevant to the Training of Psychology Students?

Students not only tended to report that they found the test highly effective (see above); they also reported that they would, on average, consider using the test as psychologists to assess the personality of one of their future patients. In total, 77% of psychology students rated the possibility that they would use the test in a professional capacity at 3 or higher (i.e., 4 or higher on the 6-point scale). The distribution of ratings is represented in Figure 2. The average rating for intention to use the test ($M = 3.17$, $SD = 1.47$) was significantly above the neutral point of 2.5, $t(1247) = 16.32$, $p < .001$, $d = 0.46$. This suggests that the test was perceived as relevant to the students' training. Of secondary interest, the degree to which students reported considering the use of the test in their future profession correlated with their perceived effectiveness for the test, $r = .62$, $p < .001$.

Figure 2

Proportion of Students in the 2019-2022 Cohorts who Would Consider Using the Fake Test with Patients



Does Experiencing the Barnum Effect Improve Engagement in the Course?

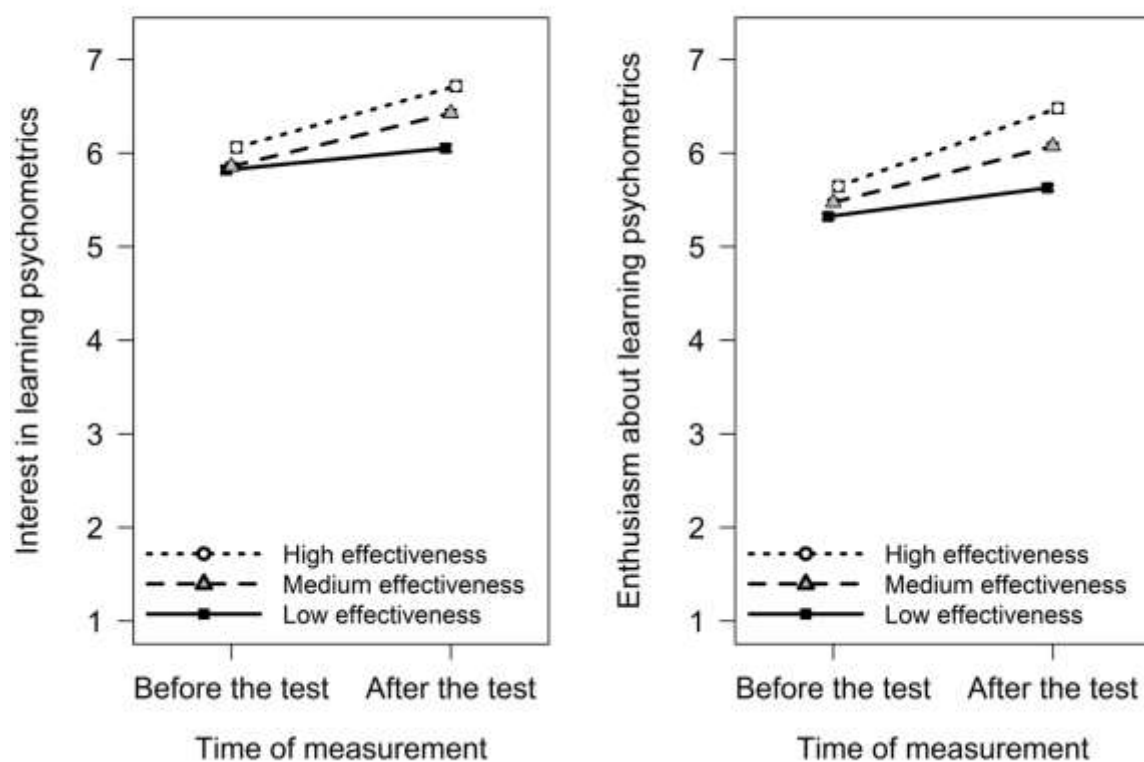
Completing the test made student perception of the course significantly more favorable. Reported enthusiasm about psychometrics significantly increased from before completing the test ($M = 5.48$, $ET = 1.46$) to after being debriefed about the Barnum effect ($M = 6.01$, $ET = 1.30$), $t(1247) = 17.66$, $p < .001$, $d_z = 0.50$. Inspecting individual profiles indicated that enthusiasm increased for 474 students and decreased for only 86 students (enthusiasm remained stable for 688 students, but 371 of these students [54%] were already at ceiling and could not have demonstrated an increase). On average, interest in psychometrics also significantly increased from before completing the test ($M = 5.88$, $SD = 1.31$) to after receiving the short debriefing about the Barnum effect ($M = 6.33$, $SD = 1.10$), $t(1247) = 13.54$, $p < .001$, $d_z = 0.38$. Interest increased for 418 students and decreased for 113 students (it remained stable for 718 students, but 523 of these students [73%] were already at ceiling).

Of secondary interest, a complementary analysis tested whether increases in interest and enthusiasm were moderated by the extent to which students had believed the test was effective (as rated after receiving their bogus personality profile) before being debriefed about the Barnum effect. Indeed, it could be the case that believing in the test to a greater extent creates a greater cognitive conflict when the deception is revealed, thus increasing interest to a greater extent; or conversely, that being deceived to a greater extent decreases students' trust in the teacher and disengages them from the course. The increase in interest was moderated by perceived effectiveness of the test, $F(1, 1246) = 37.10$, $p < .001$, $\eta^2_p = .03$, and the same was true for the increase in enthusiasm, $F(1, 1246) = 52.08$, $p < .001$, $\eta^2_p = .04$. In both cases, students who perceived the fake test as more effective demonstrated a greater increase in interest and enthusiasm about learning psychometrics when they were debriefed about the Barnum effect. These interactions are displayed in Figure 3 (for ease of

representation, perceived effectiveness of the test was recoded as a categorical variable, with an effectiveness rating from 0 to 3 recoded as "low", a rating equal to 4 (the median) recoded as "medium", and a rating equal to 5 recoded as "high effectiveness").

Figure 3

Increase in Interest and Enthusiasm About Learning Psychometrics After Experiencing the Barnum Effect, as a Function of Perceived Effectiveness for the Fake Test



Note. Perceived effectiveness was recoded as a categorical variable, for this figure only. Error bars represent within-subjects standard errors of the mean (Morey, 2008).

Does Experiencing the Barnum Effect Relate to Exam Grades?

Performance on the final psychometrics exam was higher in students who completed the fake personality test ($M = 12.38$ out of 20, $SD = 4.74$, $n = 511$) than in students who did not ($M = 11.03$ out of 20, $SD = 5.49$, $n = 265$), and the difference was significant, $t(774) = 3.56$, $p < .001$, $d = 0.27$. The results were similar when balancing group sizes by

downsampling the group of students who took the fake test. The effect was still significant in an ANCOVA controlling for grades on two other courses (cognitive psychology and experimental methodology and statistics), $F(1, 627) = 10.22, p = .001, \eta^2_p = .02$, indicating that this was not due solely to a sampling bias related to self-selection of participants choosing to participate in the procedure.

Discussion

Over seven decades after its first description by Forer (1949), the Barnum effect still appears to work well, even when implemented with a short automatic online test. In a sample of over 2000 psychology students, 85% of students exposed to the fake personality test rated its effectiveness above the neutral point, and 77% of students even rated their intention to use the test in their professional capacity as future psychologists above the neutral point, making the exercise particularly relevant to teaching in a faculty of psychology. Students not only tended to believe in the fake test: on average, directly experiencing the Barnum effect significantly increased their engagement in the course, as reflected in both their interest in psychometrics and the degree to which they were enthusiastic about following a psychometrics course. This increase appeared for most students, especially those who had believed in their bogus profile in the first place.

Teaching Implications

As proposed by Forer (1949), experiencing the Barnum effect appears to be a successful strategy to increase student interest in psychometrics. In line with the literature, the Barnum effect seems to be highly reliable, appearing for most students even over consecutive years of use. Critically, the procedure described here requires only the time needed to put the fake test and bogus profile online, one e-mail per year to invite students to complete the test, and ideally a short discussion in class. All these steps can be based on materials available online (see Gonthier, 2024). This makes it a very low-cost and low-effort

way to introduce a course touching on the topic of psychological measurement, even for very large classes. As a result, teachers of introductory courses could consider having all students experience the Barnum effect, whether they teach a psychometrics course or an introductory course on another subject that relies heavily on psychological tests (e.g. personality).

Although not part of our research questions, our results revealed something unexpected to us: the average rating for interest about psychometrics, prior to experiencing the Barnum effect, was 5.88 out of a 7-point scale. This may partly reflect sampling bias (given that the data were collected on a voluntary basis in the context of a course on psychometrics), but this figure does mean that contrary to our own preconceptions as teachers, many students were, at baseline, rather interested in the topic of psychological testing and test design. In other words, if students tend to disengage from psychometrics courses, it is not due to students not being intrinsically interested in the subject.

Part of the disengagement that is sometimes observed may instead stem from the abstract nature of many of the topics covered in psychometrics. This disengagement can be mitigated by illustrating the importance of psychometrics in practice (e.g., by presenting excerpts from test manuals and showing how difficult they are to understand without a basic training in psychometrics, or by discussing poorly conceived psychological tests and their clinical consequences), and by having students experience things for themselves. The ideal way to have students experience psychometric concepts for themselves may well be to have them construct and assess the metric qualities of their own psychological test (see Chappell et al., 2021), but when this is unfeasible, directly experiencing the effect that provides a justification for the very existence of the field appears to be a promising alternative.

Ethical Implications

Because students are given a bogus description of their personality by a trusted source, the present teaching device raises a question regarding the ethics of deception. In the

perspective of ethical teaching, the rules of conduct that apply to psychology researchers extend to their roles as teachers (Handelsman & Woody, 2015; Landrum & McCarthy, 2012), and the attitude of teachers in relation to students is a component of teaching ethics to students, at least as much as courses about ethics in psychology (Kitchener, 1992). This makes ethical implications an important point to consider.

In research, general principles are that harmless deception is justified when there is no alternative, when it provides a benefit, and when participants are debriefed as soon as possible (APA, 2017; Boynton et al., 2013; Christensen, 1988; Smith & Richardson, 1983). Authors have devoted less attention to the ethics of deception in teaching than in research, but the same principles generally apply (Handelsman & Woody, 2015; Swenson & McCarthy, 2012). In our case, deception is unavoidable to create the expected cognitive conflict, there is a direct expected pedagogical benefit for participants, and students are debriefed immediately after having read their fake personality profile, which appears as much in line with recommendations as possible. In the perspective of teaching ethical conduct to students through model teacher behavior (Kitchener, 1992), it could be worthwhile to make this rationale explicit for students during the first course immediately following the fake test.

Assuming that deception is justified in this case, what is the effect on student well-being? Some research has suggested that participants do not always resent being deceived and can even describe the experience as enjoyable and educational, although this is not always the case (for a review, see Hertwig & Hortmann, 2008). In the present study, the fact that interest in psychometrics tended to increase more for students who had strongly believed in their bogus profile in the first place tends to support the conclusion that most students were not adversely affected by the experience. However, this does not rule out the possibility that *some* students were emotionally impacted by the deception.

To verify whether this was the case, we asked 4 additional questions to students who completed the fake test in the 2023-2024 cohort (not included in the main analyses). Out of 217 respondents, 2 students (0.9%) found it problematic that a teacher deceived them for pedagogical purposes; 4 students (1.8%) reported having felt negative emotions such as anger, sadness, or disappointment upon being informed of the deception; and 217 students (100%) felt that the fake test was a useful teaching device and that it should still be used for future courses. In other words, even the 1-2% of students who felt there was a negative side to the procedure supported its continued use. These data confirm the lack of major ethical challenges to the use of the Barnum effect as a systematic teaching device, at least in the French cultural context.

Limitations and Implications for Future Research

One limitation of the current results is the impossibility with our design to determine whether experiencing the Barnum effect increased actual outcomes of student participation in the course. French university courses do not typically control attendance or engagement with course materials, and this would be difficult to do given the number of students involved. It is thus possible that the increase in student interest was limited in time or in practical benefits. Our own strategy to capitalize on this experience was to include an extended debriefing about the Barnum effect during the first course, to discuss the reasons for vulnerability to the effect and especially its practical consequences for clinical psychologists, with the hope that it would help anchor student motivation for the rest of the course.

Students who chose to participate in the procedure and who experienced the Barnum effect ultimately demonstrated higher grades, which could tentatively reflect a benefit to academic performance. There is much less support for this conclusion than for the observed increase in course interest, however, as this pattern could also reflect a sampling bias. Indeed, it could also be the case that students who chose to participate in the procedure were in fact

more motivated to do well on the course in the first place, leading them to perform better on the exam. The statistical analysis showed that students who experienced the effect did better on the exam even when controlling for overall performance on other courses, showing that they were not just "better students"; but this pattern could still reflect differences of engagement specific to the psychometric course.

Due to this non-experimental design, the actual causal effect of the procedure on grades in our dataset is impossible to determine. There was no way around this issue in the current context: randomly assigning students to experience the Barnum effect or not would not be ethically or legally possible in our country due to the requirement that all students receive the same course and enjoy the same access to pedagogical resources (for ethical considerations, see also Gurung, 2012). In short, without collecting data from students who refused to participate in the procedure, we have no way to confidently determine the direct causal effect on grades. Future studies with a control group could help answer this question.

However, perhaps a better direction for future research would be to understand why the effect could have a positive impact on student engagement in the first place. One interpretation is that experiencing the Barnum effect creates a cognitive conflict challenging student preconception (Dreyfus et al., 1990; Hewson & Hewson, 1984; Nussbaum & Novick, 1982; Posner et al., 1982), but there are other possibilities, such as improved perception of the teacher due to the use of a non-standard teaching method. Confirming whether experiencing a cognitive conflict is indeed the mechanism by which experiencing the Barnum effect translates into increased interest in the course would be helpful and potentially relevant to designing teaching devices on other topics.

References

- American Psychological Association. (2017). *Accreditation process*. <https://irp.cdn-website.com/a14f9462/files/uploaded/Table%20%20Discipline-specific%20Knowledge.docx>
- American Psychological Association. (2023). *Ethical principles of psychologists and code of conduct*. <https://www.apa.org/ethics/code/>
- Bachrach, A. J., & Pattishall Jr., E. G. (1960) An experiment in universal and personal validation. *Psychiatry*, 23(3), 267-270.
<https://doi.org/10.1080/00332747.1960.11023225>
- Baillargeon, J., & Danis, C. (1984). Barnum meets the computer: A critical test. *Journal of Personality Assessment*, 48(4), 415–419. https://doi.org/10.1207/s15327752jpa4804_15
- Beins, B. C. (1993). Using the Barnum effect to teach about ethics and deception in research. *Teaching of Psychology*, 20(1), 33–35. https://doi.org/10.1207/s15328023top2001_6
- Boyce, T. E., & Geller, E. S. (2002). Using the Barnum effect to teach psychological research methods. *Teaching of Psychology*, 29(4), 316–318.
- Boynton, M. H., Portnoy, D. B., & Johnson, B. T. (2013). Exploring the ethics and psychological impact of deception in psychological research. *IRB: Ethics & Human Research*, 35(2), 7–13.
- British Psychological Society (2019). *Standards for the accreditation of undergraduate, conversion and integrated Masters programmes in psychology*.
<https://cms.bps.org.uk/sites/default/files/2022-07/Undergraduate%20Accreditation%20Handbook%202019.pdf>
- Chappell, C., Tomcho, T. J., & Foels, R. (2021). Content of measurement courses in the undergraduate psychology curriculum. *Scholarship of Teaching and Learning in Psychology*. <http://doi.org/10.1037/stl0000311>

Christensen, L. (1988). Deception in psychological research: When is its use justified?

Personality and Social Psychology Bulletin, 14(4), 664–675.

<https://doi.org/10.1177/0146167288144002>

Dana, R. H., & Graham, D. E. (1976). Feedback of client-relevant information and clinical practice. *Journal of Personality Assessment*, 40(5), 464–469.

https://doi.org/10.1207/s15327752jpa4005_3

Dickson, D. H., & Kelly, I. W. (1985). The “Barnum effect” in personality assessment: A review of the literature. *Psychological Reports*, 57(2), 367–382.

<https://doi.org/10.2466/pr0.1985.57.2.367>

Dies, R. R. (1972). Personal gullibility or pseudodiagnosis: A further test of the “fallacy of personal validation.” *Journal of Clinical Psychology*, 28(1), 47–50.

[https://doi.org/10.1002/1097-4679\(197201\)28:1](https://doi.org/10.1002/1097-4679(197201)28:1)

Dreyfus, A., Jungwirth, E., & Eliovitch, R. (1990). Applying the “cognitive conflict” strategy for conceptual change—some implications, difficulties, and problems. *Science Education*, 74(5), 555–569.

<https://doi.org/10.1002/sce.3730740506>

Forer, B. R. (1948). A diagnostic interest blank. *Rorschach Research Exchange*, 12, 119–129.

Forer, B. R. (1949). The fallacy of personal validation: a classroom demonstration of gullibility. *The Journal of Abnormal and Social Psychology*, 44(1), 118–123.

<https://doi.org/10.1037/h0059240>

Forer, B. R. (1968). Personal validation and the person. *Psychological Reports*, 23(3f), 1214–1214. <https://doi.org/10.2466/pr0.1968.23.3f.1214>

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59–109.

<https://doi.org/10.3102/00346543074001059>

- Furnham, A., & Schofield, S. (1987). Accepting personality test feedback: A review of the Barnum effect. *Current Psychological Research & Reviews*, 6(2), 162–178.
<https://doi.org/10.1007/BF02686623>
- Gonthier, C. (2024, March 4). *Getting students interested in psychological measurement by experiencing the Barnum effect*. Open Science Framework. <https://osf.io/sc6a4/>
- Greene, R. L. (1977). Student acceptance of generalized personality interpretations: A reexamination. *Journal of Consulting and Clinical Psychology*, 45(5), 965–966.
<https://doi.org/10.1037/0022-006X.45.5.965>
- Greene, R. L., Harris, M. E., & Macon, R. S. (1979). Another look at personal validation. *Journal of Personality Assessment*, 43(4), 419–423.
https://doi.org/10.1207/s15327752jpa4304_16
- Gurung, R. A. R. (2012). Consuming scholarship of teaching and learning: Using evidence-based pedagogy ethically. In R. E. Landrum & M. A. McCarthy (Eds.), *Teaching ethically: Challenges and opportunities* (pp. 67–76). American Psychological Association. <https://doi.org/10.1037/13496-006>
- Halperin, K., Snyder, C. R., Shenkel, R. J., & Houston, B. K. (1976). Effects of source status and message favorability on acceptance of personality feedback. *Journal of Applied Psychology*, 61(1), 85–88. <https://doi.org/10.1037/0021-9010.61.1.85>
- Handelsman, M. M., & Woody, W. D. (2015). Ethical teaching in the classroom and beyond. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education*. (pp. 113–128). Oxford University Press.
- Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, 18(1), 59–92.
<https://doi.org/10.1080/10508420701712990>

- Hewson, P.W., & Hewson, M. G. (1984). The role of conceptual conflict in conceptual change and the design of science instruction. *Instructional Science, 13*, 1-13.
<https://doi.org/10.1007/bf00051837>
- Hinrichsen, J. J., & Bradley, L. A. (1974). Situational determinants of personal validation of general personality interpretations. *Journal of Personality Assessment, 38*(6), 530–534.
<https://doi.org/10.1080/00223891.1974.10120012>
- Kitchener, K. S. (1992). Psychologist as teacher and mentor: Affirming ethical values throughout the curriculum. *Professional Psychology: Research and Practice, 23*(3), 190–195. <https://doi.org/10.1037/0735-7028.23.3.190>
- Landrum, R. E., & McCarthy, M. A. (Eds.). (2012). Teaching ethically: Challenges and opportunities. *American Psychological Association*. <https://doi.org/10.1037/13496-000>
- Meehl, P. E. (1956). Wanted - a good cookbook. *American Psychologist, 11*(6), 263–272.
<https://doi.org/10.1037/h0044164>
- Merrens, M. R., & Richards, W. S. (1970). Acceptance of generalized versus “bona fide” personality interpretation. *Psychological Reports, 27*(3), 691–694.
<https://doi.org/10.2466/pr0.1970.27.3.691>
- Merrens, M. R., & Richards, W. S. (1973). Length of personality inventory and the evaluation of a generalized personality interpretation. *Journal of Personality Assessment, 37*(1), 83–85. <https://doi.org/10.1080/00223891.1973.10119833>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 8*, 61-64.
- Nussbaum, J., & Novick, S. (1982). Alternative frameworks, conceptual conflict, and accommodation: Toward a principled teaching strategy. *Instructional Science, 11*, 183-200. <https://doi.org/10.1007/BF00414279>

Orpen, C., & Jamotte, A. (1975). The acceptance of generalized personality interpretations.

The Journal of Social Psychology, 96(1), 147–148.

<https://doi.org/10.1080/00224545.1975.9923276>

Poškus, M. S. (2014). A new way of looking at the Barnum effect and its links to personality traits in groups receiving different types of personality feedback. *Psichologija, 50*, 95–105.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227. <https://doi.org/10.1002/sce.3730660207>

Smith, S. S., & Richardson, D. (1983). Amelioration of deception and harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology, 44*(5), 1075–1082. <https://doi.org/10.1037/0022-3514.44.5.1075>

Snyder, C. R. (1974). Acceptance of personality interpretations as a function of assessment procedures. *Journal of Consulting and Clinical Psychology, 42*(1), 150.

<https://doi.org/10.1037/h0036033>

Snyder, C. R., & Larson, G. R. (1972). A further look at student acceptance of general personality interpretations. *Journal of Consulting and Clinical Psychology, 38*(3), 384–388. <https://doi.org/10.1037/h0032899>

Snyder, C. R., & Shenkel, R. J. (1976). Effects of “favorability,” modality, and relevance on acceptance of general personality interpretations prior to and after receiving diagnostic feedback. *Journal of Consulting and Clinical Psychology, 44*(1), 34–41.

<https://doi.org/10.1037/0022-006X.44.1.34>

Sundberg, N. D. (1955). The acceptability of “fake” versus “bona fide” personality test interpretations. *The Journal of Abnormal and Social Psychology, 50*(1), 145–147.

<https://doi.org/10.1037/h0044374>

Swenson, E. V., & McCarthy, M. A. (2012). Ethically conducting the scholarship of teaching and learning research. In R. E. Landrum & M. A. McCarthy (Eds.), *Teaching ethically: Challenges and opportunities* (pp. 21–29). American Psychological Association.

<https://doi.org/10.1037/13496-002>

Weinberger, L. J., & Bradley, L. A. (1980). Effects of “favorability” and type of assessment device upon acceptance of general personality interpretations. *Journal of Personality Assessment*, 44(1), 44–47. https://doi.org/10.1207/s15327752jpa4401_6