# Behavior analytic jargon does not seem to influence treatment acceptability ratings

Matthew P. Normand and Hailey E. Donohue

University of the Pacific

Some have argued that behavior analysts have insulated themselves by eschewing the vernacular and adopting idiosyncratic and sometimes counterintuitive technical terms to describe their science and practice. Because of this, behavior analysis plays a minor role in psychology and related fields and effective behavior-change interventions go unused. All told, findings about the effects of behavior-analytic jargon are mixed. Studies that provided technical terms independent of context have produced unfavorable results, whereas studies that have provided context have produced positive or neutral results, overall. This study evaluated the effects of behavioral jargon on the acceptability ratings of several applied behavior analysis interventions described in terms of varying target behaviors, populations, and settings. We presented brief vignettes adapted from published research articles that were described in either jargon or nonjargon versions. There were no appreciable differences in the rated acceptability of interventions described with or without jargon.

*Key words:*   jargon, social validity, terminology, treatment acceptability

Behavior analysis seems to have a marketing problem, perhaps because the language behavior analysts use is off-putting (Becirevic et al., 2016; Critchfield, Doepke, et al., 2017). By eschewing the vernacular and adopting idiosyncratic and sometimes counterintuitive (Lattal & Poling, 1981) technical terms to describe their science and practice, behavior analysts might have insulated themselves from the broader community. Because of this, behavior analysis plays a minor role in psychology and related fields and, perhaps more importantly, several authors have argued that behavior analytic jargon has resulted in behavior change interventions going unused (Bailey, 1991; Critchfield, Becirevic, & Reed, 2017; Critchfield & Doepke, 2018; Critchfield, Doepke, et al., 2017; Doughty et al., 2012; Foxx, 1996; Freedman, 2016; Witt

et al., 1984). However, there is scant empirical evidence for these assertions. Some or all of them might be true, but it seems better to know if they are before making wholesale changes to the longstanding workings of an entire field of research and practice.

It does seem that behavior analysis is not widely embraced, but it also is not widely known. It could be that there are too few people who encounter behavior analysis, not that many do and dislike it. Or there could be other factors that lead people to avoid it, independent of how it is described. And this is important to know, because adopting the vernacular, or even the jargon of other sciences, is not without risk. The languages of all natural sciences are characterized by jargon, and for good reason (Normand, 2019). Their use was shaped by contingencies of reinforcement involving aspects of the world observed under imperfect and changing conditions. And, absent the checks and balances of the laboratory, the controlling variables of the vernacular change over time in ways that can be problematic. Metaphorical and figurative uses of words and phrases intrude. Jargon both creates and solves problems, but abandoning scientific jargon might create

more problems than are solved. Although technical jargon might obscure the relevance and importance of behavior analysis to nonspecialists, it also enables behavior analysts to communicate in ways that enhance relevant discriminations and generalizations and lead to more effective behavior in terms of the prediction and control of behavior. It seems important to know the nature and extent of the problems created by our technical jargon before offering solutions.

Perhaps the strongest evidence for the potential negative effect that technical jargon could have on the field of behavior analysis comes from Critchfield and colleagues, who, in a series of papers (Becirevic et al., 2016; Critchfield, Becirevic, & Reed, 2017; Critchfield & Doepke, 2018; Critchfield, Doepke, et al., 2017), compared the emotional valence of technical behavior-analytic terms to those common to general science and clinical practice. Overall, the terms commonly used by behavior analysts were rated less favorably as compared to terms common to other sciences and those in the vernacular. However, Critchfield and colleagues extracted their data from an existing word corpus compiled by other researchers who presented individual words to participants to rate (Critchfield, Becirevic, & Reed, 2017; Critchfield, Doepke, et al., 2017), or they presented selected technical intervention terms absent any description of the intervention (Becirevic et al., 2016). Hence, the words were evaluated outside of any syntactical arrangement, and most certainly out of the context of behavior analytic uses.

This lack of context could have important implications for how we should interpret these findings. Overall, the empirical literature is mixed on the question of how much our words matter. Several early investigations (Kazdin & Cole, 1981; Rhoades & Kratochwill, 1992; Woolfolk et al., 1977) suggested that behavior-analytic jargon had a neutral or positive effect on reported treatment acceptability. For example, Kazdin and Cole (1981) found that the content of the interventions described

influenced the acceptability ratings reported by college students, whereas calling a treatment "behavior modification" did not, and using behavioral jargon was actually associated with more positive ratings. Similarly, Rhoades and Kratochwill (1992) reported that behavioral jargon was associated with more positive ratings from teachers who viewed videos of staged interactions between teachers and a school psychologist. Interestingly, this effect was observed when the school psychologist was shown to dictate the prescribed treatment without involving the teacher in the decision; otherwise, jargon did not influence acceptability ratings.

But the picture has not always been so rosy. Rolider et al. (1998) found that the general public responded more favorably to descriptions of behavioral interventions using conversational language, and even more favorably when the conversational description was supplemented by a statement about the potential benefits of the intervention. More recently, however, Jarmolowicz et al. (2008) used the Rolider et al. methodology and found that technical language did not negatively influence treatment preference reported by direct care staff of an inpatient treatment facility for young children with intellectual disabilities. But technical language did negatively influence treatment integrity during subsequent observations of the same staff implementing the described interventions (Jarmolowicz et al., 2008). Banks et al. (2018) reported similar findings from parents who watched video descriptions of a time-out intervention described using technical, nontechnical, or "popular terminology." The parents did not prefer any one description over another, but they did understand and remember nontechnical descriptions more than technical or popular descriptions. Several other studies (Witt et al., 1984; Woolfolk et al., 1977; Woolfolk & Woolfolk, 1979) also have reported mixed results. Witt et al. (1984) found that teachers preferred pragmatic descriptions of behavioral interventions over both behavioral and humanistic descriptions,

though the effect was less pronounced when the behavior problem described was severe.

Woolfolk and Woolfolk (1979) reported that behavioral language was associated with less favorable ratings by both undergraduate and graduate students who read descriptions of classroom procedures described in "behavior science terminology," although softening the language did not consistently improve ratings. Woolfolk and Woolfolk (1979) was a systematic replication of an earlier study in which students viewed a brief video of an actual teacher using a reinforcement-based intervention in a classroom. These earlier findings (Woolfolk et al., 1977) are especially compelling because they had all participants view the same video, and the only thing that varied was that they labeled the video shown to one group of participants "behavior modification" and the video shown to a second group "humanistic education." Still, these findings do not say much about jargon, insofar as the intervention procedures were demonstrated, not described, and the only description used were the labels for the procedures (similar to Becirevic et al., 2016).

All told, there is a mixed bag of evidence on the matter of jargon and the reported acceptability of behavior-analytic interventions. Only two studies that provided context for behavior jargon reported negative effects of the jargon (Rolider et al., 1998; Woolfolk & Woolfolk, 1979). However, many of the aforementioned studies are decades old, and much can change in terms of the practices of our verbal communities across even shorter timespans. Whether any of these findings apply today is unknown. In one of the few recent evaluations of the acceptability of behavioral jargon in clinical contexts, Shemanski et al. (2018) had approximately 100 elementary school teachers rate two versions of a single vignette describing a basic behavioral intervention to decrease out-of-seat behavior exhibited by a student in a third-grade classroom. They found no differences in the rated acceptability of a behavioral intervention when it was described in jargon and nonjargon terms. This squares with

the earlier studies reporting no obvious negative influences of behavioral jargon (Kazdin & Cole, 1981; Rhoades & Kratochwill, 1992; Woolfolk & Woolfolk, 1979).

To date, research on the effect of behavior analytic jargon has focused primarily on behavior analysis in practice. Several studies have included professionals, instead of or in addition to consumers, but the professionals have been limited to behavior therapists (e.g., Rolider et al., 1998) who presumably have a bias toward behavior analytic interventions, direct care staff (e.g., Jarmolowicz et al., 2008), and elementary school teachers (e.g., Rhoades & Kratochwill, 1992). In light of this, the current study was a systematic replication of Becirevic et al. (2016), insofar as we evaluated the effects of behavioral jargon on the acceptability ratings of applied behavior analysis interventions by surveying the general population. However, like some previous studies (e.g., Shemanski et al., 2018), we presented brief vignettes, rather than just technical and nontechnical terms, described in either jargon or nonjargon versions. Unlike previous studies, the jargon versions were adapted from published research articles, using language that was the same or similar to what appeared in the original articles. The nonjargon versions were revised versions of the jargon with all technical language replaced with words from the vernacular with similar meanings. Additionally, unlike most previous studies (but see Rolider et al., 1998 for an exception), multiple vignettes were presented that depicted scenarios across a range of settings, populations, and target behaviors.

## Method

### Participants

We recruited 150 participants using the Amazon Mechanical Turk (mTurk; https://www.mturk.com) crowdsourcing marketplace, with 141 respondents (94%) returning complete surveys. All participants were 18 years of age or older and lived in the United States. (See Figure 1 and

"Demographic Characteristics" below for more information.) Participants received a $1.50 mTurk participant reward for submitting the survey, which took approximately 10 min to complete.

## Measures and Materials

We delivered the online survey using the Qualtrics cloud-based Experience Management survey software platform (https://www.qualtrics.com). We created two versions for each of eight total vignettes describing hypothetical cases of challenging behavior and subsequent hypothetical proposed treatments (see an example in the Appendix at the end of this document and the Online Supplemental Materials for all other terms). The vignettes used by Shemanski et al. (2018) served as models for the basic arrangement of those used in the present study, with each vignette consisting of a brief description of the individual and the challenging behavior, followed by a brief description of the proposed intervention. The vignettes spanned different settings

(e.g., classroom, business and industry), target behaviors (e.g., disruption, physical activity), and interventions (e.g., functional communication training, time-out), as described in Table 1. Unlike Shemanski et al., the background and proposed treatments presented in the present vignettes were adapted from research articles published in *Journal of Applied Behavior Analysis* (see Table 1 for citations to the articles used).

The jargon descriptions contained the language from the sourced research articles, as we assumed that articles published in the flagship applied behavior analysis research journal would be representative of authors describing interventions in technical jargon. To create the nonjargon version of each vignette, we replaced, wherever possible, a technical behavior-analytic term or phrase with a less-technical synonym. After all the vignettes (jargon and nonjargon) were drafted, they were independently reviewed by an informal panel of four Board Certified Behavior Analysts. The panel members provided feedback about how accurate and representative each vignette was, in terms of

**Figure 1**

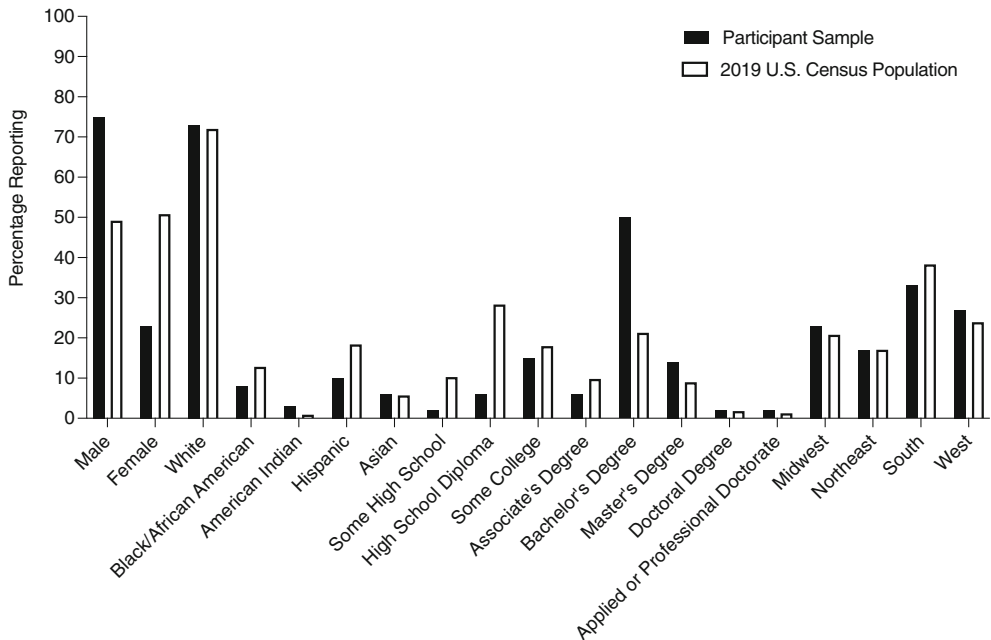*Participant Demographics Compared to the U. S. Population*

**Table 1**

*Vignette Setting, Intervention Type, and Peer-Reviewed Source*

| Population | Setting | Target behavior(s) | Intervention | Source |
|---|---|---|---|---|
| Student | Preschool | Disruption, aggression | Time-out | Donaldson and Vollmer (2011) |
| Adult | Not specified | Physical activity | Contingency management | Kurti and Dallery (2013) |
| Employee | Workplace | Safety behavior | Response effort | Abellon and Wilder (2014) |
| Adult diagnosed with dementia | Not specified | Aggression | Noncontingent reinforcement | Baker et al. (2006) |
| Adolescent diagnosed with autism spectrum disorder | Day program | Tantrums, aggression, elopement | Functional communication training | Carr and Durand (1985) |
| Student diagnosed with developmental disabilities | Elementary school | Disruption, aggression | Motivating operation | McComas et al. (2003) |
| Student | Preschool | Disruption | Stimulus control | Tiger and Hanley (2004) |
| Child | Not specified | Physical activity | Variable-ratio reinforcement | DeLuca and Holborn (1992) |

content and style, with minor revisions made based on this feedback.

## Procedure

After answering a brief series of demographics questions, each participant responded to a survey containing all eight vignettes, with four vignettes presented in the jargon versions and four in the nonjargon versions. Two groups were created for the Qualtrics survey so that participants could be randomly assigned to view four of the vignettes in the jargon versions and the other four in the nonjargon versions. We also randomized the order in which each participant viewed the vignettes. All survey responses were anonymous, and no identifying information was collected at any point during the study.

Following each vignette, participants completed a modified version of the Treatment Evaluation Inventory-Short Form (TEI-SF; Kelley et al., 1989), which was used to measure the self-reported acceptability of the intervention description. The language in the TEI-SF was modified to address the acceptability of interventions for individuals of all ages, as opposed to just children. The TEI-SF comprised nine items, each rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). An instructed response item (Appendix A, item g) was included in the survey to screen for careless responding. This item instructed participants to respond "Strongly Disagree" to questions 1-4 and "Strongly Agree" to questions 5-9 for that item.

## Results

### Demographic Characteristics

Figure 1 summarizes the demographic characteristics of the respondents compared to the most recent U.S. Census data (United States Census Bureau, 2019a; United States Census Bureau, 2019b; United States Census Bureau, 2019c). Overall, the respondents were representative of the larger U.S population.[1] Sixty-four

---

[1] We should note that although the demographic characteristics of the respondents approximate those of the U.S. population (see Figure 1), it is not clear they are a representative sample for present purposes. To start, regular users of Amazon mTurk differ in some important ways from the general population (Hitlin, 2016; Walters et al., 2018). The primary concern about the way we talk probably concerns potential consumers of our science and practice, as well as scientists and other professionals in fields for which behavior analysis should be relevant.

participants responded accurately to the instructed response item; the other 77 responses were excluded from the data analysis. The remaining 64 participants ranged in age from 25-69 years, with a mean of 35 years and a mode of 30 years. When asked in a free-response format to specify how they currently describe their gender identity, 48 (75%) participants indicated that they identified as male, 15 (23%) as female, and one respondent indicated that they did not understand the question. Thirty-five (55%) participants indicated that they are or have been a parent or guardian.

### Health

Fifteen (24%) participants indicated that they have been diagnosed with a disability or impairment; of those 15 participants, 10 indicated they had received a mental health diagnosis, four indicated that they had received a mobility impairment diagnosis, three indicated that they had received a learning disability diagnosis, and three indicated that they had received a sensory impairment diagnosis.

### Race and Ethnicity

When asked to select all categories that describe them, 46 (73%) participants indicated that "white (e.g., German, Irish, English, Italian, Polish, or French)" described them. Five (8%) participants indicated that "Black or African American (e.g., Jamaican, Haitian, Nigerian, Ethiopian, or Somalian)" described them. Two (3%) participants indicated that "American Indian or Alaska Native (e.g., Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community)" described them. Six (10%) participants indicated that "Hispanic, Latino, or Spanish Origin (e.g., Mexican or Mexican American, Puerto Rican, Cuban, Salvadoran, Dominican, Columbian)" described them. Four (6%) participants indicated that "Asian (e.g., Chinese, Filipino, Asian Indian, Vietnamese, Korean, or Japanese)" described them.

### Employment

When asked in a multiple-choice format to indicate their current employment status, 50 (78%) participants indicated that they were currently employed full-time (i.e., 40 hr or more per week). Seven (11%) participants indicated that they were currently employed part-time (i.e., fewer than 40 hr per week). Four (6%) participants indicated that they were currently unemployed. Two (3%) participants indicated that they were retired, and one (2%) participant indicated that they were a student.

### Education

One participant (2%) indicated they had completed some high school, six (9%) participants indicated they had received their high school diploma, 11 (17%) participants indicated they had completed some college, four (6%) participants indicated they had received an associate's degree, 32 (50%) participants indicated they had received a bachelor's degree, nine (14%) participants indicated they had received a master's degree, two (3%) participants indicated they had received a doctorate, one (2%) participant indicated they had received an applied or professional doctorate degree (e.g., MD, DDC, DDS, JD, or PharmD), and one (2%) participant indicated they had undergone vocational training. In response to separate questions, 35 (55%) participants indicated that they had taken a course in psychology, and 19 (30%) participants indicated that they had taken a course in behavioral psychology.

### Region

When asked to indicate in which part of the U.S. they currently reside, 15 (23%) participants indicated they currently resided in the Midwest (IL, IN, IA, KS, MI, MN, MO, NE, OH, ND, SD, WI), 11 (17%) participants indicated they resided in the Northeast (CT, MN, MA, NH, NJ, NY, PA, RI, VT),

21 (33%) participants indicated they resided in the South (AR, AL, DE, DC, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV), and 17 (27%) participants indicated they resided in the West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY). No participants indicated that they currently resided in Puerto Rico or other U.S. territories, or any other region not previously listed.

## Treatment Acceptability

Seven of the eight total vignettes are depicted in Figure 2, as these respondents all answered the eighth vignette (the instructed response item) in the same way, as instructed. The highest possible score on the TEI-SF was 45, and the lowest possible score was 9, with higher scores indicating greater acceptability. A score of 27 represented the most neutral rating. Visual analysis of the mean acceptability scores on the TEI-SF reveals no obvious differences between the jargon and nonjargon versions of the vignettes, or among any of the vignettes, irrespective of form. Acceptability scores across all vignettes, jargon and nonjargon, ranged from 28.47 to 35.89, meaning that, in aggregate, all eight vignettes were rated acceptable (i.e., ratings higher than 27), whether they contained jargon or not. Figure 2 depicts the mean TEI-SF scores and 95% confidence intervals by vignette. The means of the TEI-SF acceptability scores by demographic characteristic are depicted in Figure 3. No differences correlated with demographic characteristics such as education or disability are evident in the TEI-SF ratings.

Although visual analysis of the mean acceptability scores on the TEI-SF revealed no obvious difference related to the various demographic characteristics, we also compared the participant ratings of the jargon and nonjargon vignettes in terms of statistical significance. We used the two-tailed Wilcoxon matched-pairs signed rank test because

**Figure 2**

*Mean Acceptability Ratings and 95% Confidence Intervals for Jargon and Nonjargon Vignettes*
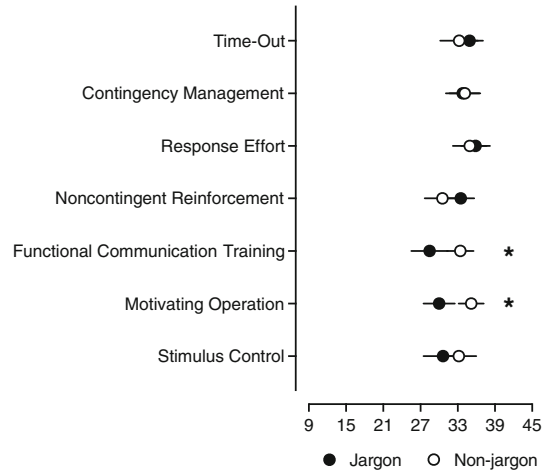


*Note.* Asterisks indicate statistically significant differences.

participant responses were not normally distributed according to the D'Agostino-Pearson omnibus normality test. We obtained statistically significant differences for only two vignettes, functional communication training

**Figure 3**

*Treatment Acceptability Scores According to Key Reported Demographic Characteristics*
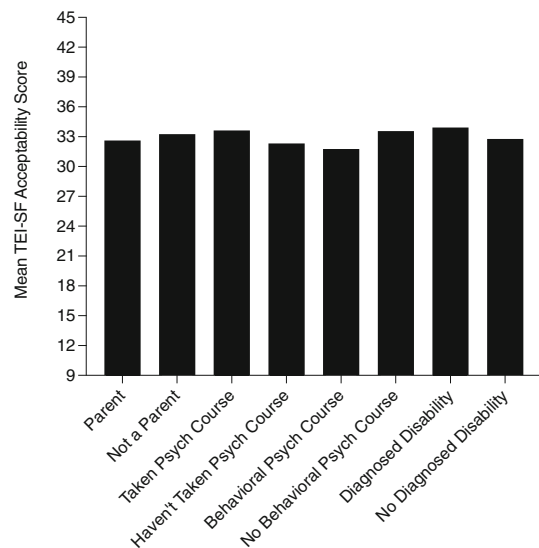
and motivating operation (see Appendix, items d and e, for descriptions).

## Discussion

The results of the current study suggest that describing behavior analytic interventions using technical jargon does not have a marked influence on the self-reported acceptability of the intervention. Only two vignettes (functional communication training and motivating operation; see Appendix items d and e) produced an observable and statistically significant difference in ratings for jargon versus nonjargon versions. But even with these differences, the ratings for all vignettes still were in the acceptable range. Overall, these findings are consistent with most of the previous research on the acceptability of behavioral jargon that did not find obviously negative effects of jargon (Jarmolowicz et al., 2008; Kazdin & Cole, 1981; Rhoades & Kratochwill, 1992; Shemanski et al., 2018; Woolfolk et al., 1977). This might mean that how we describe our interventions will not meaningfully influence the likelihood that those interventions will be adopted or recommended. It might also mean, more generally, that the way we talk about our science and practice will not meaningfully influence its adoption or use, but we cannot be certain.

For one thing, it is important to note that what people say often differs from what they otherwise do. And because someone reports liking one thing better than another does not necessarily mean they will do that one thing rather than the other. When it comes to medical procedures or car repairs, we are sometimes left to choose the less pleasing option because the decision is influenced by factors other than the emotional valence of the described procedure. But this also cuts the other way. Even if people report that they find a treatment acceptable, and many people do (Schwartz & Baer, 1991), they might not choose it when the circumstance arises. So even if the terms used to describe behavior analytic practices are less preferred, the practices might still be selected. For example, Critchfield, Doepke, et al. (2017) reported that the word *punishment* was rated both unpleasant and unmotivating. However, Hanley et al. (2005) demonstrated that some consumers chose to experience interventions that involved punishment procedures over those that did not, a somewhat counterintuitive finding.

This study did not ask the participants to choose an intervention or to choose among different descriptions of the same intervention. Instead, each participant saw either a jargon or a nonjargon version of each vignette. They might have selected the nonjargon version more often if we had presented both versions to each participant. However, few consumers of behavior-analytic practice are likely to make choices this way. Instead, many consumers probably are referred to or seek out a provider and then are presented a single type of treatment to consider. Their choice is based more on the available information influencing the selection of the provider, which might not have much to do with technical descriptions of interventions. Unfortunately, most people in need of behavioral services probably do not have the luxury of getting multiple opinions, and there might be few providers from which to choose.

On the other hand, scientists and other professionals in psychology and related disciplines are faced with a concurrent-operants situation in which they can read a variety of studies or about a variety of perspectives on a given topic, with some of what they read evoking relatively more commiserative verbal behavior.[2] And some of what they read probably allows them

---

[2]The same could be true of stakeholders and the general public in terms of access to commentary about behavior analysis on social media and on the internet (blogs, vlogs, opinion essays, etc.), more generally.

to behave more effectively, even if it is only with respect to their social and professional interactions with their colleagues. Talking and writing about B. F. Skinner and schedules of reinforcement is probably less likely to be reinforced than talking and writing about Kahneman and Tversky and decision-making. In such circumstances, the way our science is described might matter, and this should be investigated in the future.

It also seems important to at least consider that the problem, if there is one, is *what* is being described by behavior analysts, not *how* it is being described (e.g., Kazdin & Cole, 1981). The conceptual foundations of behavior analysis are anathema to much of society because they run counter to the prevailing practices of our verbal communities. These conceptual foundations focus on the circumstantial causes of behavior and, hence, require changing those circumstances to change behavior (see Friman, 2021). As Witt and Martens (1983) noted,

> Teachers' decisions to adopt behavioral interventions may be based not primarily on their effectiveness, but instead on pragmatics. For example, some interventions are too time consuming, some are excessively complicated, a few require extra personnel, and some require materials such as reinforcers that are too expensive to be used on a regular basis. For these reasons, some interventions may be simply *unacceptable* to classroom teachers. (p. 510)

And some evidence suggests that interventions are viewed more favorably when the problem behavior is severe, and that positive procedures are considered more acceptable than restrictive procedures for mild, as opposed to severe, problem behavior (Miltenberger, 1990). It is not clear, then, why the jargon versions of the FCT and MO vignettes in the current study produced somewhat lower ratings than the nonjargon versions. Both the FCT and the MO vignettes described a procedure to reduce problem behavior (disruption, aggression), and approximately the same number and kind of technical terms (e.g., contingent, reinforcement, response, discriminative stimulus) were used across the jargon versions of all the vignettes. Moreover, several of the other vignettes described similar problem behavior and produced no observable differences in ratings across the jargon and nonjargon versions.

Even when the language used to describe an intervention is softened, the behavior-analytic intervention might still be less preferred than others (e.g., Woolfolk & Woolfolk, 1979). In this study, we described the same intervention in different terms and saw no notable differences in acceptability ratings. A more thorough analysis of the effects of behavior analytic jargon should probably compare descriptions of behavior analytic interventions, with and without jargon, to descriptions of other interventions, with and without jargon, that differ along important dimensions such as the response effort required of caregivers, the frequency or duration of the intervention, and so on. We need to know more about the many variables that could influence intervention preference and adherence (e.g., Allen & Warzak, 2000; Chadwell et al., 2018; Schwartz & Baer, 1991).

Of course, it is quite possible that technical language does negatively influence intervention acceptability and the jargon versions of the vignettes in our study were not sufficiently jargon laden to detect this. However, we were careful to use the language contained in the source articles when writing the jargon versions of the vignettes. If the language used in the published scientific literature is not sufficiently jargon laden, it is difficult to see what language would be. Some behavior analysts in practice, or those who teach, might speak using more jargon-laden language, but those verbal sources are more difficult to capture. Still, future research might do well to include samples of verbal behavior taken from more diverse

sources (e.g., conference presentations, online instructional videos).

Having said all of this, we do not disagree with the notion that how we talk about our science and practice can matter. Moreover, we think the issue of technical jargon warrants multiple lines of research that address the problem in multiple ways, perhaps even some research that treats the individual word as the unit of analysis (e.g., Becirevic et al., 2016; Critchfield, Becirevic, & Reed, 2017; Critchfield & Doepke, 2018; Critchfield, Doepke, et al., 2017). But at the end of the day, how someone "feels" about a word or phrase is less important than how effectively they behave in response to it. We might not like brushing our teeth, but most of us do it frequently and do it reasonably well. Exercise is a different story. Presumably, the problem with exercise, as opposed to teeth brushing, is not meaningfully related to how these two different activities are described. As mentioned above, one important difference might be the effort involved in completing the act. This seems to be something worth investigating.

A foundational premise of behavior analysis is that behavior occurs in context. It seems prudent, then, to be cautious about making too much of our verbal behavior independent of the context in which it occurs. Presumably, a word like *love*, in isolation, would evoke positive emotional reactions. Many phrases containing the word love, such as *I would love to see you tonight*, also would evoke positive emotional reactions. But other phrases containing love, such as *I would love for you to lose your job*, would likely evoke negative emotional reactions. Some researchers (e.g., Becirevic et al., 2016; Critchfield, Becirevic, & Reed, 2017; Critchfield, Doepke, et al., 2017) have noted the relevance of context, but argue that individual words still are a valid unit of analysis for studying the emotional valence of our jargon. We disagree, because the preponderance of evidence suggests that providing context changes what people say about behavior-analytic interventions insofar as the studies that

provided context reported neutral (Kazdin & Cole, 1981; Rhoades & Kratochwill, 1992; Woolfolk et al., 1977) or equivocal (Witt et al., 1984; Woolfolk et al., 1977; Woolfolk & Woolfolk, 1979) effects. In contrast, the studies that did not provide context are the ones that have reported unfavorable effects (Becirevic et al., 2016; Critchfield, Becirevic, & Reed, 2017; Critchfield & Doepke, 2018; Critchfield, Doepke, et al., 2017). Context seems to matter.

## REFERENCES

Abellon, O. E., & Wilder, D. A. (2014). The effect of equipment proximity on safe performance in a manufacturing setting. *Journal of Applied Behavior Analysis*, *47*(3), 628-632. https://doi.org/10.1002/jaba.137

Allen, K. D., & Warzak, W. J. (2000). The problem of parental nonadherence in clinical behavior analysis: Effective treatment is not enough. *Journal of Applied Behavior Analysis*, *33*(3), 373-391. doi.org/10.1901%2Fjaba.2000.33-373

Bailey, J. S. (1991). Marketing behavior analysis requires different talk. *Journal of Applied Behavior Analysis*, *24*(3), 445-448. https://doi.org/10.1901/jaba.1991.24-445

Baker, J. C., Hanley, G. P., & Matthews, R. M. (2006). Staff-administered functional analysis and treatment of aggression by an elder with dementia. *Journal of Applied Behavior Analysis*, *39*(4), 469-474. https://doi.org/10.1901/jaba.2006.80-05

Banks, B. M., Shriver, M. D., Chadwell, M. R., & Allen, K. D. (2018). An examination of behavioral treatment wording on acceptability and understanding. *Behavioral Interventions*, *33*(3), 260-270. https://doi.org/10.1002/bin.1521

Becirevic, A., Critchfield, T. S., & Reed, D. D. (2016). On the social acceptability of behavior-analytic terms: Crowdsourced comparisons of lay and technical language. *The Behavior Analyst*, *39*(2), 305-317. https://doi.org/10.1007/s40614-016-0067-4

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis*, *18*(2), 111-126.

Chadwell, M. R., Sikorski, J. D., Roberts, H., & Allen, K. D. (2018). Process versus content in delivering ABA services: Does process matter when you have content that works? *Behavior Analysis: Research and Practice*, *19*(1), 14-22. https://doi.org/10.1037/bar0000143

Critchfield, T. S., Becirevic, A., & Reed, D. D. (2017). On the social validity of behavior-analytic

communication: A call for research and description of one method. *The Analysis of Verbal Behavior*, 33(1), 1-23. https://doi.org/10.1007/s40616-017-0077-7

Critchfield, T. S., & Doepke, K. J. (2018). Emotional overtones of behavior analysis terms in English and five other languages. *Behavior Analysis in Practice*, 11(2), 97-105. https://doi.org/10.1007/s40617-018-0222-3

Critchfield, T. S., Doepke, K. J., Kimberly Epting, L., Becirevic, A., Reed, D. D., Fienup, D. M., Kremsreiter, J. L., & Ecott, C. L. (2017). Normative emotional responses to behavior analysis jargon or how not to use words to win friends and influence people. *Behavior Analysis in Practice*, 10(2), 97-106. https://doi.org/10.1007/s40617-016-0161-9

DeLuca, R. V., & Holborn, S. W. (1992). Effects of a variable-ratio reinforcement schedule with changing criteria on exercise in obese and non-obese boys. *Journal of Applied Behavior Analysis*, 25(3), 671-679. 10.1901%2Fjaba.1992.25-671

Donaldson, J. M., & Vollmer, T. R. (2011). An evaluation and comparison of time-out procedures with and without release contingencies. *Journal of Applied Behavior Analysis*, 44(4), 693-705. https://doi.org/10.1901/jaba.2011.44-693

Doughty, A. H., Holloway, C., Shields, M. C., & Kennedy, L. E. (2012). Marketing behavior analysis requires (really) different talk: A critique of Kohn (2005) and a(nother) call to arms. *Behavior and Social Issues*, 21(1), 115-134. https://doi.org/10.5210/bsi.v21i0.3914

Foxx, R. M. (1996). Translating the covenant: The behavior analyst as ambassador and translator. *The Behavior Analyst*, 19(2), 147-161. https://doi.org/10.1007/BF03393162

Freedman, D. H. (2016). Improving public perception of behavior analysis. *The Behavior Analyst*, 39(1), 89-95. https://doi.org/10.1007/s40614-015-0045-2

Friman, P. C. (2021). There is no such thing as a bad boy: The circumstances view of problem behavior. *Journal of Applied Behavior Analysis*, 54(2), 636-653. https://doi.org/10.1002/jaba.816

Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of Applied Behavior Analysis*, 38(1), 51-65. https://doi.org/10.1901/jaba.2005.6-04

Hitlin, P. (2016). *Research in the crowdsourcing age: A case study* [Report]. P. R. Center. https://apo.org.au/node/206386

Jarmolowicz, D. P., Kahng, S., Ingvarsson, E. T., Goysovich, R., Heggemeyer, R., Gregory, M. K., & Taylor, S. J. (2008). Effects of conversational versus technical language on treatment preference and integrity. *Intellectual and Developmental Disabilities*, 46(3), 190-199. https://doi.org/10.1352/2008.46:190-199

Kazdin, A. E., & Cole, P. M. (1981). Attitudes and labeling biases toward behavior modification: The effects of labels, content, and jargon. *Behavior Therapy*, 12(1), 56-68. https://doi.org/10.1016/S0005-7894(81)80106-2

Kelley, M. L., Heffer, R. W., Gresham, F. M., & Elliott, S. N. (1989). Development of a modified treatment evaluation inventory. *Journal of Psychopathology and Behavioral Assessment*, 11(3), 235-247. https://doi.org/10.1007/BF00960495

Kurti, A. N., & Dallery, J. (2013). Internet-based contingency management increases walking in sedentary adults. *Journal of Applied Behavior Analysis*, 46(3), 568-581. https://doi.org/10.1002/jaba.58

Lattal, K. A., & Poling, A. D. (1981). Describing response-event relations: Babel revisited. *The Behavior Analyst*, 4(2), 143-152. https://doi.org/10.1007/BF03391861

McComas, J. J., Thompson, A., & Johnson, L. (2003). The effects of presession attention on problem behavior maintained by different reinforcers. *Journal of Applied Behavior Analysis*, 36(3), 297-307. https://doi.org/10.1901/jaba.2003.36-297

Miltenberger, R. G. (1990). Assessment of treatment acceptability: A review of the literature. *Topics in Early Childhood Education*, 10(3), 24-38. https://doi.org/10.1177/027112149001000304

Normand, M. P. (2019). The language of science. *Perspectives on Behavior Science*, 42(3), 675-688. https://doi.org/10.1007/s40614-017-0123-8

Rhoades, M. M., & Kratochwill, T. R. (1992). Teacher reactions to behavioral consultation: An analysis of language and involvement. *School Psychology Quarterly*, 7(1), 47-59. https://doi.org/10.1037/h0088247

Rolider, A., Axelrod, S., & Houten, R. V. (1998). Don't speak behaviorism to me: How to clearly and effectively communicate behavioral interventions to the general public. *Child & Family Behavior Therapy*, 20(2), 39-56. https://doi.org/10.1300/J019v20n02_03

Schwartz, I. S., & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis*, 24(2), 189-204. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1279564/pdf/jaba00020-0015.pdf

Shemanski, K., DeMatteo, F. J., & Fruehan, A. (2018). Behavior jargon and teacher acceptability of positive behavior interventions. *School Psychology International*, 39(5), 543-556. https://doi.org/10.1177/0143034318794747

Tiger, J. H., & Hanley, G. P. (2004). Developing stimulus control of preschooler mands: An analysis of schedule-correlated and contingency-specifying stimuli. *Journal of Applied Behavior Analysis*, 37(4), 517-521. https://doi.org/10.1901/jaba.2004.37-517

United States Census Bureau (2019a). *Table 1. Annual estimates of the resident population for the United States, regions, states, and Puerto Rico: April 1, 2010 to July 1, 2019.* [Data set]. https://census.gov/newsroom/press-kits/2019/national-state-estimates.html

United States Census Bureau (2019b). *ACS demographic and housing estimates, 2019: ACS 1-year estimates* [Data set]. https://data.census.gov/cedsci/table?q=United%20States&tid=ACSDP1Y2019.DP05&hidePreview=false

United States Census Bureau (2019c). *Table 1. Educational attainment of the population 18 years and over, by age, sex, race, and Hispanic origin: 2019, all races*. [Data set]. https://www.census.gov/data/tables/2019/demo/educational-attainment/cps-detailed-tables.html

Walters, K., Christakis, D. A., & Wright, D. R. (2018). Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? *PLoS ONE*, *13*(6), e0198835. https://doi.org/10.1371/journal.pone.0198835

Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools*, *20*(4), 510-517. https://doi.org/10.1002/1520-6807(198310)20:4

Witt, J. C., Moe, G., Gutkin, T. B., & Andrews, L. (1984). The effect of saying the same thing in different ways: The problem of language and jargon in school-based consultation. *Journal of School Psychology*, *22*(4), 361-367. https://doi.org/10.1016/0022-4405(84)90023-2

Woolfolk, A. E., Woolfolk, R. L., & Wilson, G. T. (1977). A rose by any other name …: Labelling bias and attitudes toward behavior modification. *Journal of Consulting and Clinical Psychology*, *45*(2), 184-191. https://doi.org/10.1037/0022-006X.45.2.184

Woolfolk, R. L., & Woolfolk, A. E. (1979). Modifying the effect of the behavior modification label. *Behavior Therapy*, *10*(4), 575-578. https://doi.org/10.1016/S0005-7894(79)80060-X

## Appendix A: Example of Jargon and Nonjargon Vignette

| Time-Out |
| --- |
| Donaldson, J. M., & Vollmer, T. R. (2011). An evaluation and comparison of time-out procedures with and without release contingencies. *Journal of Applied Behavior Analysis, 44*(4), 693–705. 10.1901/jaba.2011.44-693 |

| Nonjargon | Jargon |
| --- | --- |
| Background | Background |
| Jose is a 4-year-old student in preschool. Jose does not follow playground rules. He disrupts others and continues to do things when he is told not to. Jose pushes other children, runs on and jumps off of playground equipment, and throws things. Jose has injured other students and himself. | Jose is a 4-year-old student in preschool. Jose does not follow playground rules. He disrupts others and continues to do things when he is told not to. Jose pushes other children, runs on and jumps off of playground equipment, and throws things. Jose has injured other students and himself. |
| Intervention | Intervention |
| One way to improve Jose's inappropriate playground behavior is to remove Jose from the playground when he behaves inappropriately. | One way to improve Jose's inappropriate playground behavior is to use time-out from positive reinforcement. In order to |

This plan will teach Jose not to engage in inappropriate behavior. In order to teach this, he will be briefly separated from the other students and removed from the area when he engages in disruptive behaviors. This will require Jose to sit on a bench away from the playground for 4 minutes after breaking playground rules. As a result, Jose's disruptive behavior on the playground should decrease.

reduce Jose's inappropriate playground behavior, there will be an immediate withdrawal from positive reinforcement when Jose engages in disruptive behaviors. This procedure will require Jose to sit on a bench away from the playground for 4 minutes contingent on problem behavior. As a result of withdrawing Jose from the reinforcing environment as a consequence of the target behavior, Jose's disruptive behavior on the playground should decrease.

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website.