

The Development of Behavioral Performance Level Descriptors



Christian A. Latino, PhD candidate

Jason Way, PhD

Alexander Colbow, PhD candidate

Samantha Bouwers, MLIS

Alex Casillas, PhD

Tamera McKinniss, PhD

ACT[®]

Christian Latino is completing his PhD at the University of Iowa while working as a research assistant at ACT. His research interests include yoga and first generation college students.

Jason Way is a research psychologist in Behavior and Navigation Research. He researches behavioral and psychosocial skills, including their relevance to important academic and work outcomes.

Alexander Colbow is completing his PhD at the University of Iowa while working as a research assistant for ACT. His research interests include the promotion of academic success, social class, telepsychology, and gender.

Samantha Bowers is a librarian in ACT's Information Resource Center. She facilitates research projects at ACT through knowledge management, literature reviews, and process documentation.

Alex Casillas is a principal research psychologist in Research on Assessment and Learning specializing in assessment design and behavioral predictors of performance and persistence in education and work settings.

Tamera McKinniss is a senior research psychologist in Behavior and Navigation Research. She is involved in the development of behavioral skill assessments, including research determining the relevance of these skill assessments to important academic and workplace outcomes..

Acknowledgments

The authors would like to thank Jill McVey, Nicole Petersen, and Ryan Whorton for their contribution to this project.

Table of Contents

The Development of Behavioral Performance Level Descriptors	1
Overview of the Behavioral Skills Framework and Performance Level Descriptors	1
Methodology for Creation of the Performance Level Descriptors	3
Stage 1: Focus Groups to Generate Behavioral Statements	3
Stage 2: Cleaning and Categorization	4
Stage 3: Content Review of Behavioral Effectiveness Statements	4
Stage 4: Survey to Validate the Relative Effectiveness of Behaviors	4
Stage 5: Statistical Analyses	5
Finalized Performance Level Descriptors	7
Conclusion	8
References	9
Appendix A	11
SME Guide for Writing Behavioral Skill Statements	11
General Tips	11
Targeted Instructions	11
Appendix B	12
Methods for Writing PLDs	12
Appendix C	13
Recruitment Screening Questions	13
Workforce supervisors	13
Postsecondary/college	14
9th–12th grade	15
6th–8th grade	15
3rd–5th grade	16
Appendix D	17
Example Analysis of the Establishment of Categories of Effectiveness	17
Appendix E	18
Example of Analysis of Effectiveness Level Mean Differences	18

List of Tables

Table 1. SMEs in the Validation Study, by Education and Workforce Segment	4
Table 2. Example of Overlapping Ranges of the Levels of Effectiveness	5
Table 3. Agreement of Classification Across SMEs	6
Table 4. Comparisons Across Effectiveness Levels Rated by SMEs	7
Table A1. Example Analysis Used to Establish Categories of Effectiveness for the Interacting with Others Subcomponent (Socializing with Others [domain], Sociability [component])	17
Table A2. Example Effectiveness Level Comparison Analysis for the Maintaining Composure domain at the Workforce Level	18
Table A3. Post-Hoc Comparisons of Effectiveness Levels	19

List of Figures

Figure 1. Nested Structure of Performance Level Descriptors	2
Figure 2. Example of PLDs from Forgiveness Subcomponent in Postsecondary	8

The Development of Behavioral Performance Level Descriptors

As part of ACT's Holistic Framework, ACT researchers developed a Behavioral Skills Framework that can be used to assess behavior in education and work settings. By utilizing specific statements of effective behaviors, known as performance level descriptors (PLDs) (e.g., communicating well in group tasks), the Behavioral Skills Framework is designed to predict successful outcomes, identify potential risks, and propose actionable resolutions by assessing non-cognitive factors (i.e., factors not related to one's intellectual capacity). The present document contains the following:

- A brief overview of the theory and scientific evidence for the Behavioral Skills Framework
- A comprehensive description of the methods used to develop the PLDs
- A description of the anticipated uses of PLDs in assessment

Overview of the Behavioral Skills Framework and Performance Level Descriptors

Personality has been identified as an important factor in both education and workplace success (Barrick & Mount, 1991; McAbee, Oswald, & Connelly, 2014; Poropat, 2009) and can—through the identification of critical behavioral expressions—be utilized to help both students and employees achieve successful outcomes in their lives. To develop an organized structure of behavior, existing models of personality (such as the Five Factor Model [FFM; Goldberg, 1990; Peabody & Goldberg, 1989]) can be utilized to identify specific domains of behavior. In this project, the HEXACO model of personality (Ashton & Lee, 2007) was used because it builds on the traditionally-used FFM.¹ Specifically, the HEXACO model was developed on cross-cultural empirical evidence (Ashton et al., 2004), includes the Honesty-Humility domain as a sixth factor of personality (Ashton & Lee, 2007), and provides specific components (or facets) of behavior beyond broad domains (Lee & Ashton, 2004). This last element is particularly important because research has demonstrated that specific components of behavior have incremental predictive validity over broad domains of behavior alone (Chamorro-Premuzic & Furnham, 2003). In some cases, they are also able to predict specific behavioral outcomes better than broad domains (Stewart, 1999).

As such, ACT's Behavioral Skills Framework consists of levels of behavior that range from the more general domain level to the more specific PLD level. These levels of behavior, from broad to specific, are domains, components, subcomponents, and PLDs (see Figure 1). PLDs are nested within a subcomponent (e.g., Collaboration²) and are the behavioral expression of that particular subcomponent (e.g., "Works well with others," which is one aspect of being a strong collaborator, is a facet of the subcomponent Collaboration). In total, the Behavioral Skills Framework contains six domains, 23 components, 50 subcomponents, and PLDs for each of the 50 subcomponents.

¹ In the Behavioral Skills Framework, the names of the six HEXACO domains were modified for ease of understanding as follows (with the HEXACO name first): Honesty-Humility = Acting Honestly, Emotionality = Maintaining Composure, Extraversion = Socializing with Others, Agreeableness = Getting Along with Others, Conscientiousness = Sustaining Effort, and Openness to Experience = Keeping an Open Mind.

² Collaboration is defined as: Completes group tasks and achieves group goals by effectively interacting with others.

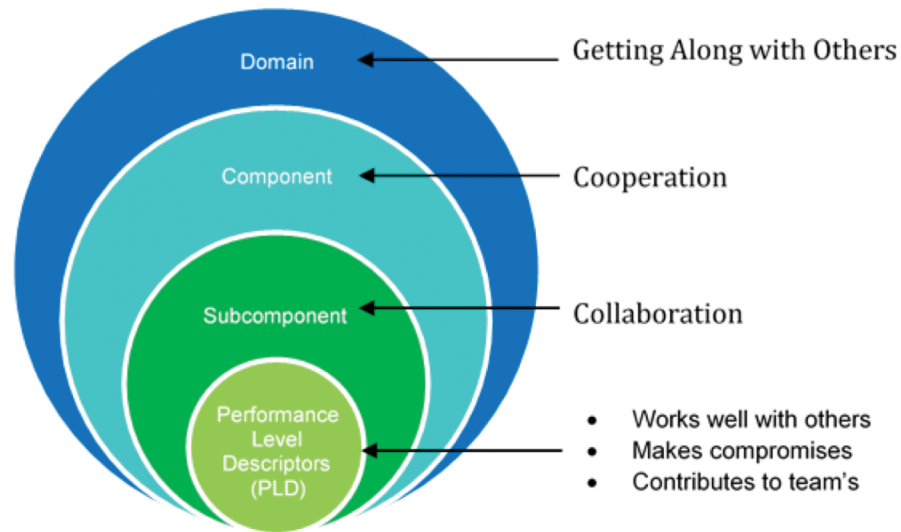


Figure 1. Nested Structure of Performance Level Descriptors

PLDs are a series of statements regarding the effectiveness of specific behavioral skills in specific settings. As found by Borman (1979) in his description of job tasks, PLDs (or “Behavior Summary Statements”) are designed to be specific enough to capture nuances of academic or work performance without being so specific that individuals have a difficult time connecting them to performance outcomes. PLDs were developed for the elementary school, middle school, high school, postsecondary, and workforce settings. Moreover, each PLD on which individuals can be rated is categorized by level of effectiveness (i.e., Highly Effective, Effective, Somewhat Effective, and Not Effective) and therefore contains multiple statements capturing different degrees of effective behavior. Each effectiveness level contains approximately two to five statements, resulting in 8 to 20 total statements that serve as the PLD of each subcomponent (see Figure 2 for an example). The purpose of these statements is to represent the content and range of behaviors within each subcomponent. For example, an individual who exhibits the subcomponent Collaboration in a *Highly Effective* manner “seeks, identifies, and engages in opportunities to work with others and expresses interest in learning from others,” while an individual who exhibits the Collaboration subcomponent in a *Not Effective* manner “only wants to pursue his/her own ideas and resists feedback.”

Ultimately, the PLDs can be employed to measure specific levels of effective individual behavior within an assessment. Different from most assessments that tend to target general areas of behavior (e.g., conscientiousness), PLDs can identify distinct and observable behaviors that can be modified (such as becoming better at sharing credit with others). In turn, this information can be used to identify areas of strength and areas of growth in behavior that can contribute to the holistic growth of the individual beyond that of an intellectual assessment. Additional information about the Behavioral Skills Framework can be found in ACT’s white paper titled *Beyond Academics: A Holistic Framework for Enhancing Education and Workplace Success* (Camara, O’Connor, Mattern, & Hanson, 2015).

Methodology for Creation of the Performance Level Descriptors

A multistage process was utilized to create the PLDs (each stage is described in greater detail in the following sections). Development of the PLDs was initiated after the establishment of the domains, components, and subcomponents of the Behavioral Skills Framework.

To define the PLDs for each subcomponent, a series of focus groups with subject matter experts (SMEs) was conducted to generate a list of behavioral statements. After cleaning and categorization, the behavioral statements were compared to existing behavioral standards to ensure comprehensiveness (i.e., that the behavioral statements captured behaviors frequently represented in behavioral standards). The SMEs then participated in a survey to rank the effectiveness of the behavioral statements for each subcomponent. Statistical analyses of the survey data were then conducted to validate the levels of effectiveness for each behavioral statement used as a PLD in the Behavioral Skills Framework.

Stage 1: Focus Groups to Generate Behavioral Statements

As advised by ACT's panel of technical experts, the development of the PLDs involved drafting an initial pool of behavioral statements during focus groups conducted with SMEs to generate a broad range of behaviors based on real-world experience. The SMEs were elementary, middle, and high school teachers; instructors and academic advisors from postsecondary institutions; and workforce supervisors (consistent with the development groups for which PLDs were developed). These particular groups of SMEs were selected because of their experience with effective (and ineffective) behaviors of students and employees in their fields. These focus groups took place remotely in two installments.

The first installment was a training session where the Behavioral Skills Framework was presented. Instructions were given for writing behavioral statements that vary by effectiveness, and SMEs practiced the statement writing task (Appendix A contains the complete instructions given to the SMEs). The second installment was an online survey where SMEs wrote statements for specific subcomponents. The SMEs were asked to generate examples of behaviors in which—relative to the expertise of the SME—students or workers would engage if they had varying levels of effectiveness for specific subcomponents. For example, the SMEs were asked to create statements of effective behavior in which individuals would engage for the Collaboration subcomponent.

The statements generated by the SMEs were reviewed by ACT researchers and were transformed into more formal PLDs (see Appendix B for methods used to write the PLDs). Specifically, the behavioral statements were transcribed into impersonal statements (i.e., in third person format), tagged with the behavior's applicable developmental group (or tagged as "general" if applicable across the lifespan), and identified as positively or negatively related to its subcomponent. Statements that aligned toward emotion or cognition were rewritten to be more behavioral in nature; if this was not possible, the statements were deleted. Next, ACT researchers reviewed the behavioral statements for the nature of the development group they belonged to, and reviewed their consistency with statements from adjacent development groups (e.g., postsecondary statements were compared to those for high school and workforce, but not middle school or elementary school). Finally, the statements were entered into a PLD database that was used to centralize the addition of new behavioral statements, track revisions, and maintain statement histories.

Stage 2: Cleaning and Categorization

The second stage of PLD development involved further cleaning and categorization of the statements in the PLD database. This included adding clarifying language, correcting grammar, and splitting statements that contained multiple behaviors into separate entries. It was also necessary to re-categorize some of the behavioral statements as subcomponent terminologies, and definitions were further refined during the collection of the behavioral statements. For example, the subcomponent Deference was changed to Cooperation, requiring a modification of behavioral statements categorized within the subcomponent.

Stage 3: Content Review of Behavioral Effectiveness Statements

After the statements generated by the SMEs were further cleaned and classified, a second and more in-depth process of nesting PLDs within subcomponents was initiated. This involved comparing the revised pool of behavioral statements from the focus groups to behavioral statements from existing assessments and standards (e.g., state standards or the College and Career Readiness standard) to ensure that critical behaviors captured in the assessments/standards were also represented in the PLD database.

Stage 4: Survey to Validate the Relative Effectiveness of Behaviors

Once the behavioral statements were finalized in the PLD database, a separate group of SMEs, each with at least three years of experience in their field, was recruited to rate the effectiveness of the behavioral statements (see Appendix C for a complete list of recruitment screening questions). As with the previous focus groups, SMEs from each of the developmental groups relevant to the Behavioral Skills Framework were recruited for this step (i.e., elementary school, middle school, high school, postsecondary, and workforce). This group was selected for their continued experience with individuals (i.e., students or employees) who have made transitions between stages of the education-to-work continuum (e.g., from middle school to high school), thus allowing them to identify a range of successful behaviors.

An online survey-hosting platform (i.e., Qualtrics) was used to assist in recruitment and survey participation. Based on previous ACT efforts to anchor ratings of student performance by teachers, an internal goal of 35 SMEs per development group was targeted for the validation of effectiveness statements. The exception to this was the workforce category for which at least 50 survey respondents were targeted to obtain a wider range of supervisors from various occupations, industries, and working styles. Table 1 displays the number of SMEs for each age group in the validation stage.

Table 1. SMEs in the Validation Study, by Education and Workforce Segment

Elementary	Middle	High	Postsecondary	Workforce	Total
35	37	37	35	53	197

The surveys were constructed to complete two tasks. First, SMEs were asked to rate the importance of each subcomponent of the Behavioral Skills Framework for success at school or in the workplace. This task was rated on a seven-point Likert-type scale (1 = Not at All Important; 7 = Extremely Important/Essential). The results of the subcomponent importance rating task are presented in a separate report. Second, SMEs were asked to rate the

effectiveness of the previously generated behavioral statements. An eight-point scale (1 = Not at All Effective; 8 = Extremely Effective) was used for this task. In order to maximize efficiency, the tasks were combined by asking SMEs to first rate the importance of a subcomponent and then rate the effectiveness of all behavioral statements (presented in random order) related to that subcomponent. Each respondent's survey data was systematically cleaned and reviewed.

Stage 5: Statistical Analyses

Once the ratings were processed, statistical analyses were conducted (See Appendix D) to establish categories of effectiveness for the statements. Based on initial evaluations (e.g., the distribution of statement ratings), it became clear that eight levels of effectiveness were too many and that raters were not able to distinguish reliably among so many levels. Instead, four levels of effectiveness (Not Effective, Somewhat Effective, Effective, Highly Effective) became the template for each developmental group (i.e., elementary school, middle school, high school, postsecondary, and workforce; see Table 2).

To review the reliability of the statements, ACT researchers used standard error of measurement (SEM) to measure consistency in scores and r_{wg} to assess interrater agreement. Higher SEM scores indicate less consistency in ratings while lower r_{wg} scores indicate less agreement in ratings among different raters. Those statements that were rated inconsistently (i.e., $SEM > .30$ and $r_{wg} < .66$) were eliminated. In a small number of cases—especially where there was initially few statements for an effectiveness level within a subcomponent—statements with r_{wg} statistics above .60 were kept as long as the SEM was below .30. Other statements were deleted based on content redundancies within an effectiveness level or when mean effectiveness ratings did not fit well into one of the four effectiveness levels. In some cases, expert judgment was also used in determining statement fit into effectiveness levels. This allowed for ACT researchers to populate as many of the levels as possible while maintaining different mean level scores for different groups of PLDs and subcomponents. Thus, the range of effectiveness ratings for each level differed slightly between each subcomponent, which is captured by the slight overlap in the ranges of effectiveness ratings in Table 2.

Table 2. Example of Overlapping Ranges of the Levels of Effectiveness

Level	Not Effective	Somewhat Effective	Effective	Highly Effective
Description	Limited skills, Least amount, Least consistent, Definitely off track	Some skills, Lesser amount, Less consistent, Just off track	Sufficient skills, Some amount, Somewhat consistent, Just on track	Most skills, Highest amount, Most consistent, Definitely on track
Ranges of Mean Effectiveness Ratings (based on 8-point scale)	1–2	2–4	5–6	6–7

SME ratings of the effectiveness levels of PLDs were analyzed in the aforementioned process to ensure that there was a high level of agreement in SMEs' perceptions of effective behavior. ACT researchers strove to only keep PLDs that were consistently viewed as being at the same level of effectiveness by large groups of educators and supervisors (i.e., the SMEs) while also being distinct from PLDs that captured different levels of effectiveness. For example, in the case of the Collaboration subcomponent, the analysis confirmed that the majority of SMEs

classified the PLD “only wants to pursue his/her own ideas and resists feedback” as *Not Effective* while the PLD “seeks, identifies, and engages in opportunities to work with others and expresses interest in learning from others” was consistently classified as *Highly Effective*.

Domain-level results are presented in Table 3 (see rows). For example, Acting Honestly had similar levels of agreement in classification across all developmental groups with the lowest level of agreement coming from workforce ($r_{wg} = .73$) and the highest level of agreement coming from middle school and high school ($r_{wg} = .81$); the overall agreement for Acting Honestly was .78. All of the domains possessed similarly high levels of agreement (range $r_{wg} = .73$ to .82; median $r_{wg} = .78$). Similarly, developmental group level results can be seen by looking at the columns of Table 3. For instance, elementary school has similar levels of agreement in classification for the six domains with the lowest level of agreement coming from Maintaining Composure ($r_{wg} = .73$), which is typically harder to observe, and the highest level of agreement coming from Socializing with Others and Sustaining Effort ($r_{wg} = .76$), which are often easier to observe. The overall agreement for elementary school was .75 with a range in the differences of agreement scores of .02. Across all domains and developmental groups, SMEs agreed 77% of the time on the effectiveness level classification of the PLDs.

Table 3. Agreement of Classification Across SMEs

	Elementary School r_{wg}	Middle School r_{wg}	High School r_{wg}	Post-secondary r_{wg}	Work-force r_{wg}	Overall r_{wg}
Acting Honestly	.75	.81	.78	.81	.73	.78
Getting Along with Others	.75	.80	.78	.81	.74	.78
Keeping an Open Mind	.74	.79	.76	.82	.75	.77
Maintaining Composure	.73	.78	.76	.80	.73	.76
Socializing with Others	.76	.80	.78	.81	.76	.78
Sustaining Effort	.76	.81	.79	.82	.75	.79
Overall	.75	.80	.78	.81	.74	.77

One-way analysis of variance (one-way ANOVA) and intraclass correlation (ICC) analyses were also conducted on the effectiveness of the behavioral statements rated by the SMEs. The one-way ANOVA can be used to determine if there are statistical differences between the means of unrelated groups while the ICC measures how similar or different items within a group are to one another. For this project, the purpose of the one-way ANOVA was to assess mean differences between the rated effectiveness levels for each set of PLDs. Specifically, it was important to ensure that the differences between the mean ratings of effectiveness levels were statistically significant (e.g., a statistically significant difference between *Somewhat Effective* and *Effective*). For these analyses, outliers were first removed from the datasets. F-values, a ratio showing how similar or different the groups means are from one another, indicated that mean ratings differed between effectiveness levels for all subcomponents in all developmental groups. However, F-values only indicate that at least one of the pairs of levels differ in the set, not which two specific levels those are. As such, to verify that adjacent levels

of effectiveness were distinguishable by raters (e.g., *Not Effective* was distinguishable from *Somewhat Effective*), Tukey's honestly significantly different (HSD) post-hoc test (Tukey, 1949) was employed to identify exactly which pairs of levels differed from each other at a statistically significant level. Statistically significant differences between effectiveness level mean ratings were found in 93.5% of all comparisons (Table 4). More specifically, demonstrating the consistency of effectiveness level differences in the Behavioral Skills Framework, 90% or greater of comparisons in each development group were identified as significantly different. For example, 96.7% of the mean ratings were significantly different within postsecondary while 96% were significantly different within workforce. This demonstrates that the vast majority of the PLDs rated by the SMEs were indeed viewed as differentiable from one another, thus truly capturing different levels of effective behavior within a subcomponent. Sample results can be found in Appendix E.

Table 4. Comparisons Across Effectiveness Levels Rated by SMEs

	Total Number of Comparisons	Number of Non-Significant Comparisons	Percentage of Significant Comparisons
Elementary School	132	8	93.9%
Middle School	144	13	91.0%
High School	150	15	90.0%
Postsecondary	150	5	96.7%
Workforce	150	6	96.0%
All Development Groups Combined	756	47	93.5%

ICCs were calculated to assess within-group agreement by developmental group across SMEs' effectiveness ratings for the set of behavioral statements within a subcomponent. For these analyses, the full datasets were used rather than the dataset with outliers removed. This decision was made because: (1) it was important to reduce the amount of missing data in the data set as it cannot be accounted for in ICC analyses; (2) a comparison between the full ICC dataset with outliers included and particular subcomponent ICC datasets with outliers removed revealed a negligible difference between the two datasets. In all cases, the ICCs were very close at > .90 (above the acceptable convention of > .80). ICCs provide useful supplemental information to the previously calculated r_{wg} statistics because they indicate the level of agreement across a group of statements rather than just for one unique statement. Sample results can be found in Appendix E.

Based on the processes described in this document, additional changes were made to wording and layout of the PLDs to be consistent with the obtained effectiveness levels for each subcomponent.

Finalized Performance Level Descriptors

The PLDs were finalized and formatted into tables, as shown in Figure 2. The mean effectiveness rating for each effectiveness level is included in parentheses after the level title (e.g., "Highly Effective (7.15)"). Consistent with the validation survey, effectiveness is rated on an eight-point scale where higher ratings indicate more effective behavior. The number of statements within each level is different for each subcomponent. This is due, in part, to the number of statements that were initially generated and retained (based both on theoretical

and statistical considerations) during the review process. Figure 2 contains an example of a PLD for individuals in postsecondary education. In addition to the varying levels of PLDs, this example illustrates the behavior framework hierarchy, going from domain (Getting Along with Others) to component (Goodwill) to the subcomponent articulated by the PLDs (Forgiveness). A comprehensive list of PLDs can be found in ACT’s technical report *ACT Behavioral Performance Level Descriptors* (Casillas, Way, McKinniss, Colbow, & Hilleman, 2016).

The envisioned purpose of the PLDs is twofold. First, PLDs can inform the development of future behavioral assessments. Specifically, PLDs—or specific observable behaviors that can lead to success in applied settings—may serve as the foundation for academic and workplace assessments that will identify areas of strength and opportunities for development at the individual level. Since PLDs are more specific than any current personality-based assessment of behavior, PLDs may aid individuals in better understanding what they must specifically change about their actions in order to achieve the desired outcome for success. For instance, rather than knowing that they simply must be more conscientious, individuals will know, concretely, where to improve their behavior within conscientiousness (e.g., improving on following through on work without reminders). Second, PLDs can inform the development of training and intervention curriculum to assist students and workers to leverage their strengths and develop areas of need. In the same way that learning progressions can inform instruction in core academic areas, PLDs can be used by curriculum experts to develop more specific and aligned instruction to support the development of more effective behaviors across education and workplace contexts. Through exposure to more specifically-tailored curriculum, students and workers will be more likely to develop the skills they need to achieve academic or workplace success.

Dimension	Getting Along well with Others (Agreeableness)
Component	Goodwill
Subcomponent	Forgiveness - Continues to work or interact with others even after others have wronged (e.g., deceived, hurt) him or her as appropriate.
Performance Level Descriptors for Forgiveness	
Highly Effective (7.15)	<ul style="list-style-type: none"> • Pleasantly works with others following conflict • Readily accepts apologies from others • Consistently gives others a second chance without holding a grudge
Effective (6.23)	<ul style="list-style-type: none"> • Often gives others a second chance without holding a grudge • Is generally willing to work with others following conflict
Somewhat Effective (3.04)	<ul style="list-style-type: none"> • Grudgingly works with others following conflict • May need convincing in order to accept apologies from others • Sometimes gives others a second chance, though he/she may continue to have hard feelings
Not Effective (1.31)	<ul style="list-style-type: none"> • Refuses to work with others following conflict • Refuses to acknowledge apologies from others • Rarely gives others a second chance and/or continues to hold grudges

Figure 2. Example of PLDs from Forgiveness Subcomponent in Postsecondary

Conclusion

Research has demonstrated that the measurement of broad personality traits can predict both academic and workforce success (Barrick & Mount, 1991; McAbee et al., 2014; Poropat, 2009). However, evidence suggests that measuring more specific behavioral expressions of

these personality traits may well improve the predictive capability of personality and behavior (Chamorro-Premuzic & Furnham, 2003; Stewart, 1999). Further, when thinking about formative assessment and training applications, it is important to have specific (diagnostic) information that can be used to provide targeted and helpful feedback. As such, ACT has developed a hierarchical framework of behavioral skills which spans from broader, general behaviors (i.e., domain-level) to very specific indicators of behavior (i.e., PLDs). ACT researchers worked with SMEs throughout education and workforce contexts to generate behavioral statements and rate their effectiveness. Statistical analyses were conducted to validate the differences in effectiveness. Findings from these analyses have established PLDs for elementary school, middle school, high school, postsecondary, and workforce settings. This spectrum of PLDs allow for a more comprehensive understanding of the qualities that make people successful in their tasks, in addition to charting how this might change over the course of an individual's life.

References

- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. <https://doi.org/10.1177/1088868306294907>
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., ... De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. <https://doi.org/10.1037/0022-3514.86.2.356>
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26.
- Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, *64*(4), 410–421. <https://doi.org/10.1037/0021-9010.64.4.410>
- Camara, W., O'Connor, R., Mattern, K., & Hanson, M. A. (Eds.). (2015). *Beyond academics: A holistic framework for enhancing education and workplace success*. Iowa City, IA, US: ACT.
- Casillas, A., Way, J., McKinniss, T., Colbow, A., & Hilleman, R. (2016). *ACT behavioral performance level descriptors*. Iowa City, IA, US: ACT.
- Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality*, *17*, 237–250. <https://doi.org/10.1002/per.473>
- Goldberg, L. R. (1990). An alternative “description of personality”: The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216>
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. http://dx.doi.org/10.1207/s15327906mbr3902_8
- McAbee, S. T., Oswald, F. L., & Connelly, B. S. (2014). Bifactor models of personality and college student performance: A broad versus narrow view. *European Journal of Personality*, *28*(6), 604–619. <https://doi.org/10.1002/per.1975>
- Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology*, *57*(3), 552–567. <https://doi.org/10.1037/0022-3514.57.3.552>

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322–338. <https://doi.org/10.1037/a0014996>

Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology*, 84(6), 959–968. <https://doi.org/10.1037/0021-9010.84.6.959>

Tukey, J. W. (1949). Comparing individual means in the Analysis of Variance. *Biometrics*, 5(2), 99–114.

Appendix A

SME Guide for Writing Behavioral Skill Statements

General Tips

1. Consider one subfacet at a time
2. Focus on behaviors that are observable
3. Keep the behavioral statements specific
4. Keep the behavioral statements simple
5. Write behavioral statements in the third person (he/she)

Targeted Instructions

1. *Think about individuals you know who display performance at each of the following levels*
Think about individuals you have taught, supervised or known who have displayed behavior at each level of performance. Focus on one performance level at a time. What does his/her behavior look like in the subfacet you are focused on?
2. *Consider different ways performance levels can vary*
The most important way is to look at the content of the specific behavioral statement. Some other ways in which performance levels could vary include (but are not limited to):
 - a. Frequency: always; sometimes; never . . . performs a specific behavior
 - b. Duration: Performs the behavior . . . For a long time; For a moderate amount of time; For a short time
 - c. Intensity: Works . . . Very hard; Somewhat hard; Not hard at all
 - d. Situational factors: Is patient . . . Even when under tight deadlines; When under moderately tight deadlines; Only when not under deadlines
3. *Write examples that differentiate between the levels of behavior*
There should be a difference between someone who is “Highly Effective” and someone “Not Effective.” For example: Patience
 - a. Highly effective: Always uses a calm speaking voice, even when he/she has a right to be angry
 - b. Not effective: Often yells when things do not go his/her way
4. *Create behavioral statements are appropriate for your age group*
 - a. Behaviors should be specific to the age group you are focused on (e.g., elementary, middle school, postsecondary, work).
 - b. Highly effective behaviors are somewhat dependent on age. Highly effective behavior in kindergarten may be considered to be typical or average once a person reaches college.
5. *Write down your statements during the live workshop on the subfacet handouts and enter them later into the online survey*

Appendix B

Methods for Writing PLDs

1. Read through all of the behavioral statements found under each subfacet. Determine which areas seem most important to the subfacet.
2. Decide which areas of behavior seem especially important to the subfacet. Highlight or take note of the behavioral statements that relate to these important areas.
3. Make a note of areas of consensus or disagreement for each level of behavior.
4. When you have determined which behavioral statements are important to the subfacet and show levels that are grade or work appropriate, you can begin to write summary statements.
5. When you have completed writing summary statements for a subfacet, double-check the following elements:
 - a. Check to be sure your summary statements are clear and distinct at each level in a way that is fairly easy to differentiate. (For example, “some of the time” and “most of the time” are distinguishable, whereas “not enough” and “enough” are likely not).
 - b. Check for grammar and spelling errors.
 - c. Make sure each level has the same number of relevant summary statements.
 - d. Make sure that all elements you identified as important are included somewhere in your summary statements.
 - e. Look over the original behavioral statements to be sure that you have not left anything important out.

Time estimates:

- Steps 1–3: Approximately 30–60 minutes per subfacet
- Step 4: Approximately 1–2 hours per subfacet
- Step 5: Approximately 5–10 minutes per subfacet

Appendix C

Recruitment Screening Questions

This appendix contains a list of screening questions for each age group. The questions are listed next to the numbers, with the answer choices listed by letter below each number. The answers that would remove the person from the survey are labeled as such.

Workforce supervisors

1. Choose the occupational field closest to yours:
 - a. Mechanical & Electrical Specialties (e.g., automotive technician, telecommunications technician)
 - b. Regulation & Protection (e.g., security guard, police officer)
 - c. Engineering & Technologies (e.g., mechanical engineer, surveyor)
 - d. Medical technologies (e.g., Optician, pharmacist)
 - e. Communications & Records (e.g., administrative assistant, hotel clerk)
 - f. Health Care (e.g., dental hygienist, medical assistant)
 - g. Manufacturing & Processing (e.g., power plant operator, welder)
 - h. Crafts & Related (e.g., chef, tailor)
 - i. Other (removes person)
2. Is supervising others a major part of your job?
 - a. Yes
 - b. No (removes person)
3. How many years of experience do you have supervising others?
 - a. 1 (removes person)
 - b. 2 (removes person)
 - c. 3
 - d. 4
 - e. 5
 - f. 6
 - g. 7
 - h. 8
 - i. 9
 - j. 10 or more

Postsecondary/college

1. From the broad categories listed below, please choose the one that best fits your job title. If none of them do, please select other:
 - a. Professor (e.g., Assistant Professor, Associate Professor)
 - b. Instructor
 - c. Lecturer
 - d. Student advisor (e.g., Associate Student Affairs Officer)
 - e. Academic advisor (e.g., Associate Academic Officer)
 - f. Other (removes person)
2. Approximately how big is your employing institution, in terms of the number of undergraduate students?
 - a. 7,500 or fewer undergraduate students
 - b. More than 7,500 undergraduate students (removes person)
3. How many years of experience do you have interacting with students one-on-one?
 - a. 1 (removes person)
 - b. 2 (removes person)
 - c. 3
 - d. 4
 - e. 5
 - f. 6
 - g. 7
 - h. 8
 - i. 9
 - j. 10 or more
4. How much of your weekly job time (in %) is spent interacting with students?
 - a. 10% (removes person)
 - b. 20% (removes person)
 - c. 30%
 - d. 40%
 - e. 50%
 - f. 60%
 - g. 70%
 - h. 80%
 - i. 90%
 - j. 100%

9th–12th grade

1. Do you have experience teaching 12th grade students?
 - a. Yes
 - b. No (removes person)
2. How many years of experience do you have teaching 12th grade students?
 - a. 1 (removes person)
 - b. 2 (removes person)
 - c. 3
 - d. 4
 - e. 5
 - f. 6
 - g. 7
 - h. 8
 - i. 9
 - j. 10 or more

6th–8th grade

1. Do you have experience teaching 9th grade students?
 - a. Yes
 - b. No (removes person)
2. How many years of experience do you have teaching 9th grade students?
 - a. 1 (removes person)
 - b. 2 (removes person)
 - c. 3
 - d. 4
 - e. 5
 - f. 6
 - g. 7
 - h. 8
 - i. 9
 - j. 10 or more

3rd–5th grade

1. Do you have experience teaching 6th grade students?
 - a. Yes
 - b. No (removes person)
2. How many years of experience do you have teaching 6th grade students?
 - a. 1 (removes person)
 - b. 2 (removes person)
 - c. 3
 - d. 4
 - e. 5
 - f. 6
 - g. 7
 - h. 8
 - i. 9
 - j. 10 or more

Appendix D

Example Analysis of the Establishment of Categories of Effectiveness

Table A1. Example Analysis Used to Establish Categories of Effectiveness for the Interacting with Others Subcomponent (Socializing with Others [domain], Sociability [component])

Item Text	Question ID	Effectiveness Level	Mean	SD	SEM	r_{wg}	n	Level
Avoids group work even when necessary	Q51_3	Not effective	1.97	1.24	0.17	0.76	37	1
Avoids interactions during group work	Q51_4	Not effective	2.20	1.26	0.18	0.76	35	1
Does not attend social activities, even when specifically invited	Q51_7	Not effective	2.44	1.30	0.18	0.75	36	1
Interacts with others as minimally as necessary during group work	Q51_12	Somewhat effective	3.17	1.59	0.22	0.70	36	2
Seeks out group work only when he/she is told to	Q51_11	Somewhat effective	4.30	1.66	0.23	0.68	37	2
Attends social events when specifically invited, but does not actively participate	Q51_10	Somewhat effective	4.84	1.71	0.24	0.67	37	2
Seeks out group work when the task requires it	Q51_5	Effective	6.72	1.16	0.16	0.78	36	3
Actively participates in social events	Q51_9	Effective	6.94	0.86	0.12	0.84	36	3
Usually interacts with all others during group work	Q51_1	Effective	6.97	1.23	0.17	0.77	36	3
Consistently makes an effort to include and interact with all others during group work	Q51_2	Highly effective	7.36	0.80	0.11	0.85	36	4
Initiates social events and actively participates	Q51_6	Highly effective	7.41	0.61	0.09	0.88	34	4
Actively seeks opportunities for group work even when the task does not always require it	Q51_8	Highly effective	6.39	1.50	0.21	0.71	36	delete

Appendix E

Example of Analysis of Effectiveness Level Mean Differences

Table A2. Example Effectiveness Level Comparison Analysis for the Maintaining Composure domain at the Workforce Level

One-Way ANOVA

Subcomponent		Sum of Squares	df	Mean Square	F	Sig.	Intraclass Correlations	Confidence Intervals (95%)	
								Lower	Upper
Worry Management	Between Groups	1207.676	3	402.559	183.549	.000	.987	.969	.997
	Within Groups	763.230	348	2.193					
	Total	1970.906	351						
Negative Feeling Management	Between Groups	1483.280	3	494.427	247.621	.000	.982	.963	.994
	Within Groups	1190.038	596	1.997					
	Total	2673.318	599						
Decisiveness	Between Groups	689.613	3	229.871	105.048	.000	.972	.938	.992
	Within Groups	986.902	451	2.188					
	Total	1676.514	454						
Independence	Between Groups	955.296	3	318.432	146.478	.000	.973	.947	.990
	Within Groups	1421.751	654	2.174					
	Total	2377.047	657						

Note. df = degrees of freedom; F = F-ratio (variation between sample means/variation within means); Sig. = p value.

Table A3. Post-Hoc Comparisons of Effectiveness Levels

Worry Management					Negative Feeling Management				
	Not Effective	Somewhat Effective	Effective	Highly Effective		Not Effective	Somewhat Effective	Effective	Highly Effective
Not Effective					Not Effective				
Somewhat Effective	x				Somewhat Effective	x			
Effective	x	x			Effective	x	x		
Highly Effective	x	x	x		Highly Effective	x	x	x	

Decisiveness					Independence				
	Not Effective	Somewhat Effective	Effective	Highly Effective		Not Effective	Somewhat Effective	Effective	Highly Effective
Not Effective					Not Effective				
Somewhat Effective	x				Somewhat Effective	x			
Effective	x	x			Effective	x	x		
Highly Effective	x	x	x		Highly Effective	x	x	x	

Note. x indicates a significant difference between two levels ($p < .05$); n.s. indicates a non-significant difference



ACT is an independent, nonprofit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year, we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.

For more information, visit www.act.org.

