Evaluating Student Outcomes from a College of Agriculture and Life Sciences Leadership Development Program

Robert Strong, Texas A&M University
Jennifer Strong, Texas A&M University
Karissa Palmer, Texas A&M University
Bailee Jo Miller, Texas A&M University

The Dr. Joe Townsend Leadership Fellows Program focuses on developing students in five specific areas, developing self, developing others, organizational management skills, vision, and values. Evaluations should assess the extent leadership programs are meeting outcomes and preparing students for post-graduate success. The purpose of this study was to determine the effectiveness of the program through the use of Kirkpatrick's evaluation model, Bloom's taxonomy, and Ajzen's theory of planned behavior. A summative evaluation was used to assess whether or not the program had been meeting its stated objectives. Students who once participated in the program since its inception was sent a survey through Qualtrics to determine the success rate of graduates through Kirkpatrick's reaction, learning, behavior, and results domains. The data collected was then analyzed to determine if the objectives of the study were met. The participants scored statistically significant scores on all four portions of Kirkpatrick's model. The overall effectiveness of the agricultural leadership program, per the data collected, was deemed successful, and the results from the participants were both positive and promising for the program. Recommendations for future practitioners and researchers, as well as other agricultural leadership programs, were delineated based on the study's data and conclusions.

Introduction/Theoretical Framework

Leadership programs offered at colleges and universities assist in the growth of future leaders and their leadership development. In the case of leader development, the emphasis typically is on individual-based knowledge, skills, and abilities associated with formal leadership roles (Day 2001). Another way to perceive leadership development is as a strategy helping people understand how to relate to others, coordinate their efforts, build commitments, and develop extended social networks by applying self-understanding to social and organizational imperatives (Strong, 2016). Leadership development, as defined by the Center for Creative Leadership, is the expansion of a person's capacity to be effective in leadership roles and processes. Leadership roles and processes are those that enable groups of people to work together in productive and meaningful ways (Van Velsor et al., 1998). Involvement in leadership programs, both before and during students' college years, is very pertinent for their growth and development as leaders. In fact, numerous post-secondary institutions are developing and formalizing the study of leadership at both undergraduate and graduate levels through majors, minors, certificates, or selected course offerings (White, 2006). Although institutions are incorporating more formal leadership learning courses, knowledge is a necessary first step, but by itself, it is not sufficient for changing leadership behavior (Kotter, 1996; Strong et al., 2013a). The new knowledge must be put into action. Skills encompass the action domain of learning (McDonald-Mann, 1998). Therefore, it is crucial for students to get involved in leadership programs (Miller, 2018; Strong & Williams, 2014). USDA's (2018) strategic plan called for increased numbers of agricultural leadership programs to develop future leaders and to maintain a competitive and stable global food supply.

Over the last 30 years, there has been a rapid increase in leadership programs at universities and colleges across North America (Dugan & Komives, 2007). However, as leadership programs continue to increase in number, a lack of consensus persists regarding how these programs should be designed to teach (Eich, 2008). Due to the increase in leadership programs, there is an even higher need for program evaluations and assessments. Not all programs are deemed effective and appropriate based on their program objectives (Lindner et al., 2016). As identified by Gall et al. (1996) educational evaluations are procedures to develop judgments about the value or merit of educational programs. To help ensure leadership programs are teaching students significant information that will help in their future leadership endeavors, program evaluations are needed to solidify the programs. Brungardt and Crawford (1996) posited assessment of leadership programs help focus program development and implementation in the needs of students. In addition to the effectiveness of leadership programs, financial support of these programs is also an important aspect to consider when looking at the value of the program. Leadership educators in an era of fiscal tightness understand the importance of program justification and survival (Brungardt & Crawford, 1996). Without the proper fiscal support, a program will not survive despite how effective the program is to the students involved. Moreover, despite the growing evidence that structured leadership programs benefit students, little is currently known about the best methods for making such interventions (Strong et al., 2013b; Strong et al., 2021). Effective leadership education must address foundational questions, such as theoretical framework, curricular content, instructional methods, and assessment (Strong et al., 2021). This is a major reason why evaluations are so important. The best way to help figure out the best methods for program development is through current program evaluations, as well as, basing future programs on the recommendations found through those assessments (Strong et al., 2021).

All levels of Kirkpatrick's (1990) evaluation model are extremely important in determining the overall effectiveness of trainings or programs. Reio et al. (2017) suggested Kirkpatrick's model is outcome and objective-oriented and focuses on determining the results of a program. Although Kirkpatrick's model does not include formative evaluation levels, its main purpose is to assess the overall effectiveness of programs. Kirkpatrick's model was not designed to evaluate the planning and creating of the leadership program, but merely to assess the final result. When using Kirkpatrick's (1990) model, it is important to assess all four levels to ensure the most accurate evaluation outcomes. In conclusion, "limiting an evaluation to one particular level almost certainly will not provide an adequate picture of the overall outcomes of any training program (Reio et al., 2017).

Kirkpatrick's (1998) model has made valuable contributions to training evaluation thinking and practice. Not only has Kirkpatrick's model been a great evaluation tool, but it also has been a building block for other evaluation models. The model has also served as a useful—if preliminary—heuristic for training evaluators (Alliger & Janak, 1989; Lee et al., 2021) and has been the seed from which a number of other evaluation models have germinated (Holton, 1996; Strong et al., 2022). The learning level of Kirkpatrick's and Kirkpatrick's (2006) model will be assisted using Bloom's taxonomy to help determine the level of learning the students achieved. In addition to Bloom's taxonomy, Ajzen's (1991) theory of planned behavior will be used to strengthen the research done on the behavior level of Kirkpatrick's model.

Created by Ajzen (1991), the theory of planned behavior actually identified a small set of causal factors that should permit explanation and prediction of most human social behaviors. Briefly, according to the theory, a central determinant of behavior is the individual's intention to perform the behavior in question (Fishbein & Ajzen, 2010). Per the theory, there are three different types of considerations that help people articulate their intentions. The connection among attitudes and behavior can be explained by an alignment to specific behavioral intentions (Bumguardner, 2014; Strong et al., 2013c). That is, attitudes about performing a behavior would predict behavioral intentions to enact the behavior, which would in turn predict behavior. Behavioral intentions are determined by attitudes towards the behavior, subjective norms surrounding the behavior, and perceived behavioral control (Ajzen, 1991). As such, according to the Theory of Planned Behavior, behavioral intentions are framed as the motivational component of the model, or one's conscious plan or decision to exert effort to perform the target behavior (Fishbein & Ajzen, 2010). The first consideration includes readily accessible or salient beliefs about the likely consequences of a contemplated course of action or performing the target behavior, beliefs which, in their aggregate, result in a favorable or unfavorable attitude toward the behavior (Fishbein & Ajzen, 2010). A second type of consideration has to do with the perceived normative expectations of relevant referent groups or individuals. Such salient normative beliefs lead to the formation of a subjective norm—the perceived social pressure to perform or not to perform the behavior (Fishbein & Ajzen, 2010). Subjective norms are "beliefs about whether others think one should engage in the behavior (Ajzen, 1991). Individuals are "assumed to take into account factors that may further or hinder their ability to perform the behavior, and these salient control beliefs lead to the formation of perceived behavioral control, which refers to the perceived capability of performing the behavior" (Fishbein & Ajzen, 2010, p. 32).

Purpose and Objectives

The purpose of this study was to assess the overall effectiveness of the Dr. Joe Townsend Leadership Fellows Program. Being a donor-sponsored program, another factor that influenced the purpose of this study was justifying its continuation to its current and future donors and university officials. The specific objectives were:

1. Describe the extent participants discerned the program achieved its objectives respective to Kirkpatrick's model.

2. Assess the level of Bloom's Taxonomy enforced through the program.

3. Define the top developed attributes, in relation to results, students strengthened while participating in the program.

4. Investigate the effects of reaction, learning, behavior, and results on program objectives.

Methodology

For this study, all 171 ($N = 171$) participants of the Dr. Joe Townsend Leadership Fellows Program planned on being surveyed and included in the evaluation. However, of the 171 participants in the program, only 108 were successfully contacted, while the other six were not reachable with the contact information we had access to. Seeing as this college program was and

still is very selective of their members, the number of students who have participated in the program are very few. With that being said, the study was most effective as a census study to get a more accurate data collection.

Conducting a census often results in enough respondents to have a high degree of statistical confidence in the survey results (Franekel et al., 2019). When it comes to program evaluations, it is imperative to give everyone the opportunity to provide feedback (Patton, 2021). The more people you have to participate in an evaluation questionnaire, the more statistically relevant your evaluation becomes. A disadvantage to census surveys is the lack of response rate. Seeing as you are trying to get everyone to respond to your survey request, not everyone is going to be able to respond and complete the survey. The most common use of a census study is through the U.S. Census Bureau.

The instrument used to conduct this study was a 24-question survey. The instrument was assessed for content validity by a team of researchers at Texas A&M University and deemed valid for the study's objectives. The evaluation contained mainly Likert-type questions, however, there was a few open-ended questions for the participants to elaborate their responses if needed (see Table 1). The questionnaire included participant personal characteristics, and the four levels of Kirkpatrick's (2006) evaluation model; reaction, learning, behavior, and results.

Table 1.
Questionnaire Questions and Section Created through Qualtrics

Survey Question Sections	Questions
Personal Demographics	The year the student participated.
	The year the student graduated.
	Age
	Race
Reaction	Overall thoughts of the program.
	Would they participate again?
	Their favorite part.
	Their least favorite part.
	What was the most impactful part of the program?
Learning	Current endeavors (student, graduate student professional, entrepreneur, other)
	Did the things they learned help their current endeavors?
	Are they holding any leadership positions?
	List and describe those leadership positions.
	Did the program help them prepare for their current endeavors?
	Did they gain any contacts through the program?
Behavior	Do they utilize the network in their current endeavors?
	Do they use any skills learned while participating in the program?
	Did their views of themselves as leaders change due to the program?
	How did being involved influence their current leadership positions?
	How well did the program help them to strengthen the Developing Self attributes?

Results	How well did the program help them to strengthen the Developing Others attributes? How well did the program help them to strengthen the Values attributes? How well did the program help them to strengthen the Organizational Management attributes? How well did the program help them to strengthen the Vision attributes?

Using the Likert-type questions and the open response boxes, the data collected was used to determine the overall success of the program. The survey was designed to help decide whether the program achieved all four stages of the Kirkpatrick's evaluation model. Although their time at Texas A&M University and how Fellows affected that time is important, the objectives of the program are directly correlated to future successes. Therefore, the elements of the program, if any, which helped students in their future endeavors will be determined.

Surveys are a remarkably useful and efficient tool for learning about people's opinions and behaviors (Dillman et al., 2014). Careful selection of survey questions through a tailored design and completing a census study helped with the data collection. Tailored Design is the development of survey procedures that work together to form the survey request and motivate various types of people to respond to the survey by establishing trust and increasing the perceived benefits of completing the survey while decreasing the expected costs of participation (Dillman et al., 2014). The survey was sent out through Qualtrics and the data was collected through Qualtrics as well. "For most surveys, for example, it is typical to receive only one maybe two contacts in a 10-day period; anymore begin to become irritating (Dillman et al., 2014). Although the "optimal timing sequence for web surveys has not, we believe, been determined yet, there are basic rules and timing sequences that should be followed (Dillman et al., 2014). When it comes to the pre-notice, the survey invitation, the thank you/reminder contact, and the final thank you email, each needs to be sent with enough time in between so that the communication is not annoying, but also so that the prior contact has not yet been forgotten (Dillman et al., 2014).

Out of the 171 students that participated in the Students program since 2009, 159 were capable of being contacted. Of the 159 contacted, 129 completed the survey and were able to be evaluated. After the initial contact was sent to the students, 71 participants completed the survey. After six days the initial contact was sent, the first reminder was sent to those participants who had yet to complete the survey. After the first reminder, 43 more participants completed the survey. Lastly, three days after the first reminder was sent, the second and final reminder was sent to those who had yet to complete the survey. After the final reminder, 15 participants completed the survey leaving our final number of respondents at 129 out of the initial 171. The response rate for the survey ended up being 81.48% (see Table 2).

Table 2.
Respondents after each Contact

Number of Respondents	f	%
After initial contact	71	44.65
After first reminder	43	27.04
After final reminder	15	11.63

Microsoft Excel and SPSS were used to analyze the data once it is all collected. By using Microsoft Excel, as well as SPSS, the respondent's answer choices were analyzed based on the five sections of the survey, personal characteristics, reaction, learning, behavior, and results. By analyzing the questions in the four sections of Kirkpatrick's model, the overall objectives of the Students program could be determined effective or not. Cronbach's alpha was utilized *ex post facto* to assess the reliability coefficients of each construct; reaction earned a .91, learning's reliability coefficient was .79, behavior scored a .84, and results earned a .83 reliability coefficient. Descriptive statistics are used to describe the basic features of the data in a study and provide simple summaries about the sample and the measures (Franekel et al., 2019). An advantage of descriptive statistics includes condensing a large amount of data into a smaller, simpler form of data. A disadvantage of descriptive statistics includes the fact that they don't go very in-depth (Franekel et al., 2019). Inferential statistics go beyond the surface of the data to find a deeper meaning of the statistics themselves (Franekel et al., 2019). The results found using Kirkpatrick's model were dependent on the other variables in the study.

When conducting the study, one of the main concerns was nonresponse error. Nonresponse error occurs when the people selected for the survey who do not respond are different from those who do respond in a way that is important to the study (Dillman et al., 2014). The best way to eliminate the possibility of a nonresponse error is to tailor the design to best meet the audience's needs. Making the survey as quickly as possible, creating an eye-appealing design, and giving good incentives as to why the respondents should participate are all ways to reduce the chance of nonresponse error (Dillman et al., 2014).

According to Lindner et al. (2001), there are three approaches to test for non-response error. The method used to test for non-response error in this study was method two. Method two consists of comparing early respondents to late respondents. Method two "is an extrapolation method in which non-respondents are a linear extension of the latest respondents, and a trend may be detected across respondents based on relative earliness or lateness to respond (Lindner et al., 2001). Early, after first contact, respondents were compared to late, second and third contact, respondents and no significance in the data existed. Therefore, nonresponse error did not exist in the data and results can be generalized to the population of Texas A&M University [Name] agricultural leadership program participants (Lindner et al, 2001).

Findings/Results

The Qualtrics survey was split into five categories that helped guide the objectives of the study using Kirkpatrick and Kirkpatrick's (2006) evaluation model. The five categories included, demographics, and the four portions of Kirkpatrick's model, reaction, learning, behavior, and results. Some of the questions pertaining to the reaction portion of Kirkpatrick's model were evaluated on a five-anchor scale, while the others were evaluated on a six-anchor scale. The learning portion was evaluated using a six-anchor scale and earned ($M = 5.24$, $SD = .03$). Behavior was evaluated using a four-anchor scale and a six-anchor scale. Lastly, the results were

evaluated using a four-anchor scale and earned (*M* = 3.32, *SD* = .06). The remaining data collected from each section of Kirkpatrick's model can be found in Table 3.

Table 3.
Four Levels of Kirkpatrick's Evaluation Model (N = 129)

Kirkpatrick's Four-Levels	*M*	*SD*
Learning	5.24	.03
Behavior	4.17	.23
Reaction	3.78	.29
Results	3.32	.06

Note. *1= Strongly Disagree, 2= Disagree, 3= Somewhat Disagree, 4= Somewhat Agree, 5= Agree, 6= Strongly Agree*

The reaction section of Kirkpatrick's evaluation model, based on a five-anchor scale, possessed questions about specific aspects of the program. The participants overall thoughts on the program were very positive (*M* = 4.18), on a scale of one to five with one being far short of expectations and five being far exceeds expectations. Of the 88 respondents, 85 selected meets expectations or higher, while only three respondents selected short of expectations. No participants selected far short of expectations. When it comes to specific aspects of the students program, the impact of guest lecturers was the highest-ranking aspect (*M* = 3.75). The lowest-ranking aspect was the impact of other aspects (*M* = 2.89). However, it is to be noted that of the rest of the aspects all participants ranked the aspect on a scale of one to five with one being least impactful and five being most impactful, while only 18 participants ranked the "other" option on the same scale (see Table 4).

Table 4.
Data from the Reactions Level of Kirkpatrick's Model (N = 129)

Reactions	*M*	*SD*
Overall thoughts on Students program	4.18	.88
Impact of guest lectures	3.75	1.44
Impact of the field trip	3.51	1.22
Impact of fellow program members	3.36	1.27
Impact of the retreat	3.07	1.27
Impact of other aspects	2.89	1.45

Note: *Overall Mean = 3.46, Overall SD = .21*

The last question analyzed through the Qualtrics survey was past participants' willingness/desire to participate in students program again if given the opportunity. This question was asked using a six-anchor scale with one being strongly disagree and six being strongly agree. Students, on average, concurred that they would participate in the Fellow program again (*M* = 5.67). The data also had a standard deviation of 0.69 (SD = 0.69). In collecting this data, all participants selected agree or higher, with only three participants selecting somewhat agree and one participant selecting disagree.

The other two questions in the reaction section were frequency questions asking what the students' favorite and least favorite parts of the Fellow's program were. Of the different parts of

the program, most participants claimed guest lecturers to be their favorite part ($f = 37$). The retreat was the option the least amount of participants selected as their favorite part ($f = 7$). When it comes to the participants' least favorite part of the Fellow's program, the least amount of Students selected the field trip ($f = 6$). The majority of the participants selected "other" as their least favorite part ($f = 59$). The respondents who selected "other" were then given the option to explain what "other" part of the program was their least favorite. Of the 59 respondents, 34 of them used the "other" option to explain that they did not describe a least favorite part. Of the remaining 25 respondents, common themes in their responses included wanting more reconnection opportunities with other alumni, wanting more time in the program and throughout the week, and the tendency for some student members to be too social during class time by getting off-topic (see Table 5).

Table 5.
Data from the Frequency Questions in the Reaction level of Kirkpatrick's Model

Kirkpatrick's Model: Reaction	*f*	%
What was your favorite part of Fellows?		
Guest Lecturers	37	42.0
The Field Trip	21	23.9
Fellow Program Members	15	17.0
Other	8	9.1
The Retreat	7	8.0
What was your least favorite part of Fellows?		
Other	59	67.0
The Retreat	12	13.6
Fellow Program Members	7	8.0
Guest Lecturers	4	4.6
The Field Trip	6	6.8

After collecting data from the reactions portion of Kirkpatrick's evaluation model, *t*-tests were used to assess whether there was a significant difference between men and women when it came to their reactions of the program. Surprisingly, the results found were not as expected. The men who participated in the program showed to have a higher mean ($M = 3.59$, SD $= .49$) than the women who participated ($M = 3.34$, SD $= .46$). Tukey's analysis revealed the effect size was small $d = .29$ (see Table 6).

Table 6.
t-test Results Comparing Reactions Scores of Male and Female (N = 129)

Reaction Scores	*n*	*M*	*SD*	*t*	*p*
Male	50	3.59	.49	2.77	.01
Female	79	3.34	.46		

Note. $p < .05$

Per the participant's responses (see Table 7), students believed that being a part of the program not only helped their current endeavors ($M = 5.25$, $SD = .87$)), but it also helped prepare them for their current endeavors as well ($M = 5.23$, $SD = .83$).

Table 7.

Data from the Learning level of Kirkpatrick's Model (N = 129)

Learning	M	SD
The things you learned helped your current endeavors	5.25	.87
Did the program help you prepare for your current endeavors	5.23	.83

Note: *1= Strongly Disagree, 2= Disagree, 3= Somewhat Disagree, 4= Somewhat Agree, 5= Agree, 6= Strongly Agree*

Note: *Overall Mean = 5.24, Overall SD = .85*

Regarding behavior per Kirkpatrick and Kirkpatrick's (2006) model, some questions were asked with a six-anchor scale, while the others were asked using a four-anchor scale. The three questions asked using a six-anchor scale can be found in Table 8 ($M = 4.95$, $SD = .14$). When asked whether being involved affected their current leadership positions, the students who participated in the survey felt that the program was very useful ($M = 5.19$, $SD = .88$). Students' views of themselves as leaders, on average, changed due to their participation in the program ($M = 4.94$, $SD = .02$). Students also agreed that they gained contacts through their involvement in the program ($M = 4.70$, $SD = .17$). Since some of the behavior level data were collected using a six-anchor scale, while the others were collected using a four-anchor scale, the six-anchor scale questions can be found in Table 8, while the four-anchor scale questions can be found in Table 9.

Table 8.

Data from the Behavior level of Kirkpatrick's Model (N = 129)

Kirkpatrick's Model: Behavior	M	SD
How did being involved in Fellows affect your current leadership positions	5.19	.88
Did your views of yourself as a leader change due to the Fellows program	4.94	.02
Did you gain any contacts through the Fellows program	4.70	.17

Note. *1= Strongly Disagree, 2= Disagree, 3= Somewhat Disagree, 4= Somewhat Agree, 5= Agree, 6= Strongly Agree*

Note. *Overall Mean= 4.95, Overall SD= .14*

Questions asked using a four-anchor scale had an average score of 3.01 ($M = 3.01$, $SD = .26$). Located in Table 10, most of the participants felt they currently use the skills they learned while participating in the program ($M = 3.60$, $SD = .56$), some participants felt they were utilizing the alumni network while others were not ($M = 2.42$, $SD = .93$). Of the 129 respondents, 53 claimed they probably or definitely utilize the alumni network, while 76 claimed they probably or definitely do not utilize the network.

Table 9.

Data from the Behavior Level of Kirkpatrick's Model (N = 129)

Behavior	M	SD
Do you use any of the skills you learned while participating in Fellows?	3.60	.56
Do you utilize the Fellows network in your current endeavors	2.42	.93

Note. *1= Strongly Disagree, 2= Disagree, 3= Agree, 4 = Strongly Agree*

Note: *Overall Mean = 3.01, Overall SD = .26*

After collecting the data from the Behavior section of Kirkpatrick's Evaluation model, *t*-tests were run to determine if there was a difference between men and women and their behavior results after participating in the Students program. Table 11 shows that the men on average scored significantly higher in the Behavior portion of Kirkpatrick's model ($M = 4.99$, $SD = .33$) in comparison to the women who participated ($M = 4.81$, $SD = .59$). The difference was significant ($p < .05$) and Tukey's analysis indicated the effect size was medium ($d = .59$).

Table 10.

t-test Results Comparing Behavior Scores of Male and Female

Behavior Scores	n	M	SD	t	p
Male	32	4.99	.33	2.01	.03
Female	54	4.81	.59		

Note. *p < .05*

Overall, of the five pillars of the program, the responses show that the Students found the greatest growth in their values ($M = 3.37$, $SD = .05$), while they found the least amount of growth in the development of their organizational management skills ($M = 3.27$, $SD = .04$). When it comes to the developing self objective, leaders are successful by learning how to be dependable, dedicated, persistent, and the rest of the attributes under the developing self pillar. The developing others objective measures success by claiming that successful leaders foster an environment where leaders and followers are able to communicate, build consensus, and other developing attributes. Organizational management means that successful leaders are able to listen, manage time, as well as the rest of the competencies. Vision involves successful leaders who are able to set goals, be creative, and other numerous attributes. Lastly, values entail successful leaders who model maturity, commitment, and fifteen other attributes also.

The five pillars were evaluated on a four-anchor scale, with one being strongly disagree and four being strongly agree. However, the means in the five pillars showed satisfying improvement and awareness. All five of the pillars have relatively similar means and standard deviations, with the entire results section being a mean score of 3.32 and standard deviation of .06 (see Table 12).

Table 11.

Overall Data from the Results level of Kirkpatrick's Model for the Five Pillars of Fellows

Results	M	SD
Values	3.37	.05
Vision	3.32	.04
Developing Others	3.31	.06
Developing Self	3.29	.06
Organizational Management Skills	3.27	.07

Note. 1= *Strongly Disagree*, 2= *Disagree*, 3 = *Agree*, 4 = *Strongly Agree*
Note: *Overall Mean = 3.32, Overall SD = .06*

Conclusions/Implications/Recommendations

The participants found the most growth in learning and results portions of Kirkpatrick's evaluation model. The other three portions of the model and survey, reaction, behavior, and results, on the other hand, were all deemed statistically significant and reliable. Although these are notable statements, growth comes from improving the areas in which the program is lacking. All of the sections of Kirkpatrick's (1990) model had very similar means, however, reaction and behavior proved to have the lowest means. These are the sections of the model that should be addressed for improvement. Surprisingly, the men who completed the survey had higher scores in both reaction and behavior than the women who completed the survey. Although this goes against most research regarding men and women in leadership programs and situations, the data found is assumed that the men had further to increase in their skills and abilities than the women.

Kirkpatrick's evaluation model has provided a straightforward system or language for talking about training outcomes and the kinds of information that can be provided to assess the extent to which training programs have achieved certain objectives (Bates, 2004). When selecting this evaluation model, it was an obvious decision to assess all four levels of Kirkpatrick's model. Although it can be common to do so, "Kirkpatrick contends that it is a serious mistake to bypass Level 1 and 2 and only conduct Level 3 and 4 evaluations (Reio et al., 2017). The problem with dismissing the importance of evaluating levels 1 and 2 is that by doing so, one "could easily lead to the wrong conclusions about the effectiveness of the intervention and the training program's overall result (Kirkpatrick & Kirkpatrick, 2006). Seeing as the purpose of this study was to determine the overall effectiveness of the Fellow's program, all four levels had to be addressed. Additionally, "limiting an evaluation to one particular level almost certainly will not provide an adequate picture of the overall outcomes of any training program (Reio et al., 2017). Based on the data collected, it is seen that "favorable reactions to training do not, by themselves, guarantee that learning (Level 2), or improved performance (Level 3) has occurred, Kirkpatrick stresses that many organizations are overlooking the importance of Level 1 evaluation (Kirkpatrick & Kirkpatrick, 2006). Kirkpatrick also "emphasizes that there can be no guarantee that a favorable reaction to the training program assures learning, positive behavioral change, and favorable organizational results (Kirkpatrick & Kirkpatrick, 2006). Although all the levels of Kirkpatrick's model were assessed, it is crucial to see the importance of Levels 1 and 2 evaluations. When comparing the collected means of men and women, it was found that the men scored higher in the reaction and behavior sections of the survey. It was also found that 58% of the participants were women and 42% were men. This type of data found is very contradictory to what most leadership studies say about men and women participants. In fact, Dillman (2014) stated as is commonly the case, women responded to the survey at higher rates than men. Although women did respond at a higher rate in this study, the percentages of men and women were much closer than anticipated.

When analyzing the data from this research study, one recommendation that can be made for future researchers is to investigate the effects of Kirkpatrick's evaluation model on the program objective using a different type of research method. Seeing as this study was a quantitative research study, to get a better idea of the personal reactions and behaviors of the Students participants, a qualitative study would be very beneficial. Qualitative research could provide insight into various problems, such as identifying some of the underlying factors that account for the weak, but statistically significant relationships sometimes found in the literature. Qualitative research may also help to identify variables that have not yet been considered or quantitatively

tested (Reio et al., 2017). When evaluating the collected data per the study's objectives, another recommendation to be made is to perform a cost-benefit analysis to help better justify the program being donor sponsored. By merely conducting data analysis to determine the effectiveness of the program, some donors may be willing to continue funding, however, others may not.

Focusing more on improving the reaction and behavior portions of Kirkpatrick's evaluation model would help the Texas A&M University's Dr. Joe Townsend Leadership Fellows Program be more efficient in meeting its objectives. However, it is important to state that all the portions of Kirkpatrick's model were relatively close in mean. However, limiting an evaluation to one particular level almost certainly will not provide an adequate picture of the overall outcomes of any training program (Reio et al., 2017). Seeing as 72.7% ($n = 64$) of the students who participated in this study claim to be currently holding leadership positions, as discussed in the demographics section, it can be inferred that participation in the program has been beneficial both during their time in the program and also in preparation for their futures. The imperative the data indicates is to focus on the least developed leadership pillars when examining for ways to increase the effectiveness of the program (Strong et al., 2021).

References

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42*, 331–342.

Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation And Program Planning, 27*(3), 341-347. https://doi.org/10.1016/j.evalprogplan.2004.04.011

Brungardt, C. & Crawford, C. B. (1996). A comprehensive approach to assessing leadership students and programs: Preliminary findings. *The Journal of Leadership Studies, 3*(1), 37–48.

Bumguardner, K. M., Strong, R., Murphrey, T. P., & Dooley, L. M. (2014). Examining the blogging habits of agricultural leadership students: Understanding motivation, use, and self-efficacy. *Journal of Agricultural Education, 55*(3), 32–42. https://doi.org/10.5032/jae.2014.03032

Day, D. V. (2001). Leadership development: A review in context. *The Leadership Quarterly Yearly Review of Leadership, 11*(4), 581–613.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Surveys: Internet, mail, and mixed-mode surveys: The Tailored Design Method* (4th ed.). Wiley & Sons.

Dugan, J. P., & Komives, S. R. (2007). *Developing leadership capacity in college students: Findings from a national study. A report from the Multi-Institutional Study of Leadership*. National Clearinghouse for Leadership programs.

Eich, D. (2008). A grounded theory of high-quality leadership programs: Perspectives from student leadership development programs in higher education. *Leadership & Organizations Studies, 15*(2), 176–187. https://doi.org/10.1177/1548051808324099

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. Psychology Press.

Fraenkel, Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education*. (10th ed.) McGraw-Hill Publishing.

Gall, M. D., Borg, W. R, & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). Longman Publishers.

Holton, E. F., III (1996). The flawed four level evaluation model. *Human Resource Development Quarterly, 11*(4), 333-360.

Icek, A. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). Berrett-Koehler Publishers

Kotter, J. P. (1996). *Leading change*. Harvard Business School Press.

Lee, C.-L., Strong, R., & Dooley, K. E. (2021). Analyzing precision agriculture adoption across the globe: A systematic review of scholarship from 1999–2020. *Sustainability, 13,* 10295. https://doi.org/10.3390/su131810295

Lindner, J. R., Rodriguez, M. T., Strong, R., Jones, D., & Layfield, D. (2016). Research priority area 2: New technologies, practices, and products adoption decisions. In Roberts, T. G., Harder, A., & Brashears, M. T. (Eds). *American Association for Agricultural Education national research agenda: 2016-2020*. Gainesville, FL: Department of Agricultural Education and Communication.

Lindner, J. R., Murphy, T., & Briers, G. E. (2001). Handling nonresponse In social science research. *Journal of Agricultural Education, 42*(4), 43–53. https://doi.org/10.5032/jae.2001.04043

McDonald-Mann, D. G. (1998). Skill-based training. In McCauley, C. D., Moxley, R. S., & Van Velsor, E. (Eds.). *The Center for Creative Leadership handbook of leadership development*. Jossey-Bass.

Miller, B. J. (2018). Utilizing the Kirkpatrick Model to Evaluate a Collegiate High-Impact Leadership Development Program. Master's thesis, Texas A&M University. https : / /hdl .handle .net /1969 .1 /173373

Patton, M. Q. (2021). Evaluation criteria for evaluating transformation: Implications for the coronavirus pandemic and the global climate emergency. *American Journal of Evaluation, 42*(1), 53–89. https://doi.org/10.1177/1098214020933689

Reio, T. J., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A critique of Kirkpatrick's Evaluation Model. *New Horizons In Adult Education & Human Resource Development, 29*(2), 35–53.

Strong, J., Strong, R. Greenberg, Purdue, E., & Dolly, D. (2016). Understanding learners as followers: A study of a US and a Caribbean University. *International Leadership Journal, 8*(2), 88–103.

Strong, R., Zoller, J., & Palmer III, J. (2022). Evaluating the Adoption of Virtual Reality Equine Selection and Judging Curricula: Instructional Responses to a COVID-19 Consequence. *Journal of International Agricultural and Extension Education, 29*(1), 76–85. https://newprairiepress.org/jiaee/vol29/iss1/6

Strong, R., Dooley, K., Murphrey, T., Strong, J., Elbert, C., & Baker, M. (2021). The EVAL framework: Developing impact evaluation scholars. *Advancements in Agricultural Development, 2*(3), 1–13. https://doi.org/10.37433/aad.v2i3.139

Strong, R., & Williams, J. (2014). Understanding students as followers discovering the influence of followership style on self-directed learning. *Journal of Agricultural Education, 55*(2), 201–213. https://doi.org/10.5032/jae.2014.02201

Strong, R., Williams, J., Irby, T. L., & Wynn, J. T. (2013a). Country club management and self-directedness: Implications for academics and practitioners of leadership. *NACTA Journal, 57*(4), 38–44. https://www.jstor.org/stable/nactajournal.57.4.38

Strong, R., & Irby, T. L., & Dooley, L. M. (2013b). Factors influencing students' behavioral intentions: Examining the potential use of mobile technology in agricultural education courses. *Journal of Agricultural Education, 54*(4), 149–161. https://doi.org/10.5032/jae.2013.04149

Strong, R., Wynn, J. T., Irby, T. L., & Lindner, J. R. (2013c). The relationship between students' leadership style and self-directed learning level. *Journal of Agricultural Education*, *54*(2), 174–185. https://doi.org/10.5032/jae.2013.02174

United States Department of Agriculture. (2018). *USDA Strategic Plan FY 2018 - 2022*. https://www.usda.gov/sites/default/files/documents/usda-strategic-plan-2018-2022.pdf

Van Velsor, E., McCauley, C. D. & Moxley, R. S. (1998). Our view of leadership development. In McCauley, C. D., Moxley, R. S., & Van Velsor, E. (Eds.). *The*

Center for Creative Leadership handbook of leadership development. Jossey-Bass.

White, B. J. (2006). Design and implementation of an Interdisciplinary Leadership Studies Minor at an Historically Black Liberal Arts College. *Journal of Leadership Education, 5*(2), 50–59.