HELPING YOUTH THRIVE

PYD RESEARCH

2023 NATIONAL 4-H INDEX STUDY






National Report

Ryan J. Gagnon
Clemson University

Mary E. Arnold
National 4-H Council

Jill Lingard
4-H Common Measures, University of Nebraska-Lincoln

Jessie Mullendore
4-H Common Measures, University of Nebraska-Lincoln

Makayla Burg
4-H Common Measures, University of Nebraska-Lincoln

NATIONAL 4-H COUNCIL

4-H COMMON MEASURES

Table of Contents

Suggested Citation

Gagnon, R. J., Arnold, M. E., Lingard, J., Mullendore, J., & Burg, M. (2023). *2023 National 4-H Annual Index Study.* Washington DC: National 4-H Council.

4-H is the youth development organization of our nation's Cooperative Extension System and USDA.

Highlights from the Study

Youth Characteristics

- 3,958 4-H youth from 19 state 4-H programs participated in the study.
- The average participant was 15.13 years old, joined 4-H at 11.08 years, predominantly identified as white (89.73%) and female (64%), had about 4.04 years of experience in 4-H, and were mostly in grades 8 to 11 (74.93%).
- The most common projects were in Animal and Agricultural Science, with a significant portion of youth engaging in at least one 4-H project (33%).
- Gender of respondents did not significantly impact their participation or involvement in 4-H activities.
- Non-white participants generally had fewer years in 4-H and were less involved in summer activities, but were more engaged at the national level.

4-H Program Outcomes

Civic Engagement

- Participants inspired to volunteer through 4-H showed greater community involvement.

College and Career Readiness

- As participants increased in age, they were more likely to report they learned about colleges that may be a good fit through 4-H.
- Older participants were also more likely to report 4-H has helped with decisions about college.
- Higher levels of volunteer inspiration correlated with increased college and career readiness and more frequent application of science in daily problems.
- Discussing science positively correlated with learning about suitable colleges.

Healthy Living

- Learning about healthier eating was linked to 4-H education on healthy food choices.

Science

- A majority of participants reported that they learned, applied, and enjoyed science through their 4-H experiences.

Demographics and Common Measures

- There were no significant statistical correlations between gender, racial identity, number of 4-H projects participated in, and the 4-H Common Measures.

Youth Thriving

- Programs that fostered a sense of belonging, supportive peer and adult relationships, and personal growth in 4-H were associated with better scores in all 4-H Common Measures areas.
- Moderate to strong connections were found between achieving developmental outcomes in 4-H (like positive academic attitudes, community connection, personal responsibility and social-emotional competence), and higher civic engagement.
- Participants experiencing youth thriving elements (like openness to challenges, curiosity, emotional intelligence, self-improvement, and goal orientation) in 4-H reported higher levels of civic engagement, college and career readiness, healthy living, and enjoyment of science.
- Youth thriving was found to partially mediate the relationship between the developmental context and outcomes, aligning with Arnold and Gagnon's 2019 findings (and the 2022 report).

2023 National 4-H Index Study Participant Information

This report presents the findings from the second annual National 4-H Index Study, examining the experiences of 4-H members. For well over a century, 4-H has contributed to the learning and development of youth in significant ways, positioning 4-H as one of the largest and most enduring youth development organizations in the country. While project-based learning and positive youth development have long been the intended outcomes of 4-H, those outcomes have not been systematically measured on a national sample. The 2023 National 4-H Index Study represents the second wave of national index data collection, investigating the relationship between 4-H program outcomes and positive youth development, and providing a point-in-time assessment of the impact of 4-H on America's youth. The results of this second wave of data collection are consistent with the first wave, indicating the stable impact of 4-H on key youth development outcomes.

All land-grant universities (LGUS) with 4-H programs were invited to participate in the study. Nineteen LGUS participated in this year's study. After the data were screened (see Appendix 1 for details) a total of 3,958 youth responses were included in the analysis. Table 1 shows the breakdown of youth participation by land-grant university.

Table 1. Youth Participation by Land-grant University

Land-Grant University	Frequency	Percent of total
Alabama	122	3.08%
California	225	5.68%
Colorado	240	6.06%
Connecticut	76	1.92%
Florida	153	3.87%
Illinois	210	5.31%
Kentucky	103	2.60%
Michigan	98	2.48%
Minnesota	139	3.51%
Montana	230	5.81%
Nebraska	276	6.97%
New Hampshire	35	0.88%
New York	59	1.49%
Northern Marianas	17	0.43%
Ohio	1,423	35.95%
Oregon	143	3.61%
Virginia	197	4.98%
West Virginia	175	4.42%
Wyoming	37	0.93%
	Total: 3,958	

Each LGU secured its own IRB approval for the study. Participants were recruited independently by each LGU, and data were submitted via a Qualtrics link to the national database managed by 4-H Common Measures. Data were collected throughout the spring and summer of 2023.

The study instrument consisted of:

- Selected items from four of the 4-H Common Measures program outcomes areas:
  1. Healthy Living
  2. Science
  3. Civic Engagement
  4. College and Career Readiness
- Items from the 4-H Thriving Model instrument that measured:
  1. The Developmental Context
  2. Youth Thriving Indicators
  3. Positive Youth Development Outcomes
- Measures of youth demographic characteristics, youth level, quality and quantity of 4-H experience

Youth Respondents Demographics

Youth ages 13 and up were invited to participate in the study. The lower age limit was established because the Thriving Model instruments are designed for youth at least 13 years of age. As shown in Table 2, respondents were an average age of 15.13 years, started in 4-H at an average age of 11.08, have been in 4-H for an average of 4.04 years, and participated in an average of 2.38 4-H projects.

Table 2. Mean Respondent Age and 4-H Participation

	Mean	SD	Min	Max	Range
Age	15.13	1.60	13	19	6
Number of Years in 4-H	4.04	1.48	0.5	5	4.5
Age at Start of 4H	11.08	1.88	8	18.5	10
Number of 4-H Projects	2.38	1.68	1	9	8

As shown in Figure 1, participants primarily self-identified as female (64%), with 33% reporting male and 1% as non-binary. Seventy-four (2%) of respondents elected not to respond to the gender identity question.

Figure 1. Respondent Gender Identity

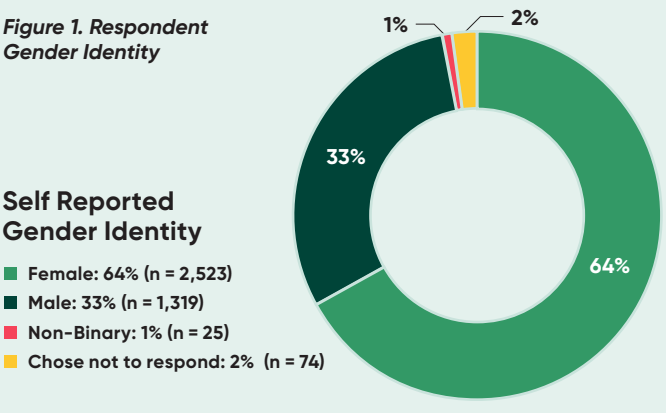


Table 3 presents the complete respondent demographic information based on youth grade, race/ethnicity, 4-H activities and number of unique 4-H projects participated in.

Table 3. Complete Youth Participation Demographics

	N	%
Grade		
6th	19	0.51%
7th	417	11.21%
8th	786	21.12%
9th	800	21.50%
10th	641	17.23%
11th	561	15.08%
12th	327	8.79%
Graduated in 2023	170	4.57%
Racial and/or Ethnic Identity		
Asian Origin	36	0.93%
African American	53	1.37%
Native American	24	0.62%
Hawaiian/Pacific Islander	5	0.13%
White (Non-Hispanic)	3,459	89.73%
Hispanic	75	1.95%
Multiple Race	203	5.27%
4-H Project Areas*		
Animal and Agricultural Science		33%
Business and Citizenship		4%
Creative Arts		16%
Diversity and Inclusion		2%
Environment & Outdoor Science		10%
Healthy Living & Foods		14%
Practical Skills		10%
Professional Development		4%
Science, Technology, Engineering & Math		8%
Total Number of 4-H Projects Participated In		
1 Unique Project	1,675	42%
2 Unique Projects	848	21%
3 Unique Projects	595	15%
4 Unique Projects	388	10%
5 Unique Projects	202	5%
6 Unique Projects	129	3%
7 Unique Projects	56	1%
8 Unique Projects	38	1%
9 Unique Projects	27	1%

*Note: *indicates a select all that apply option, Ns excluded for parsimony (and to avoid confusion on overall sample size)*

4-H Project Participation

Respondents were asked to select the projects they participated in across nine unique offerings. As illustrated in Figure 2 and Table 3 "Animal and Agricultural Science" was the most popular selection (33%). As illustrated in Figure 3 and Table 3, respondents participated in an average of 2.38 4-H projects (*SD* = 1.68 range = 1 to 9 projects), with the primacy of youth reporting participation in one unique project (*n* = 1,675, 42%). Put simply, Animal and Agricultural Science represented the largest proportion of responses to the question "what 4-H projects have you participated in?" and the primacy of youth indicated they participated in one 4-H project (42%).

Figure 2. Percentage of Youth Participation by Number of 4-H Projects

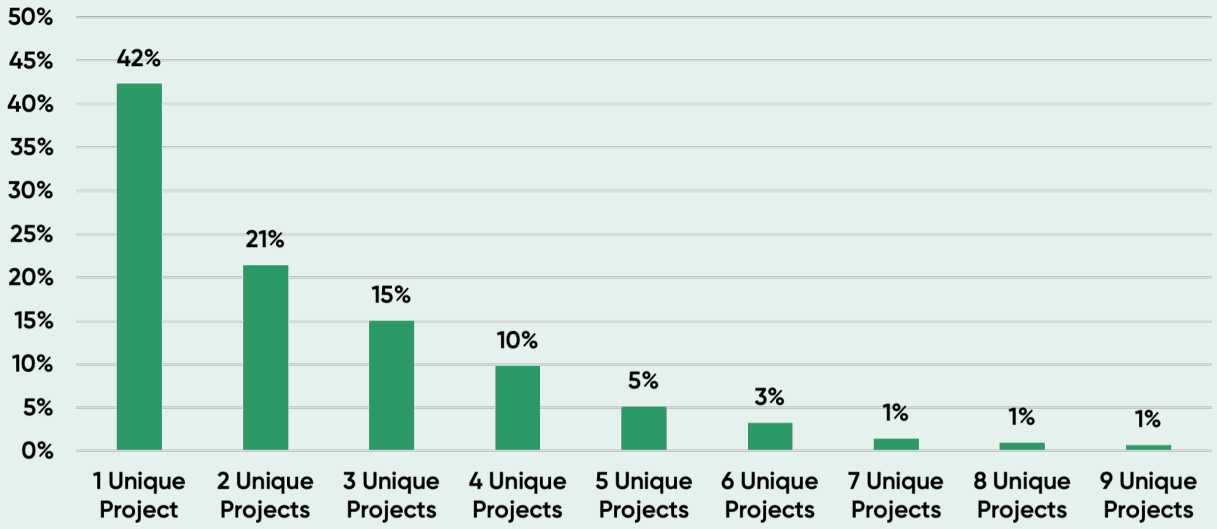
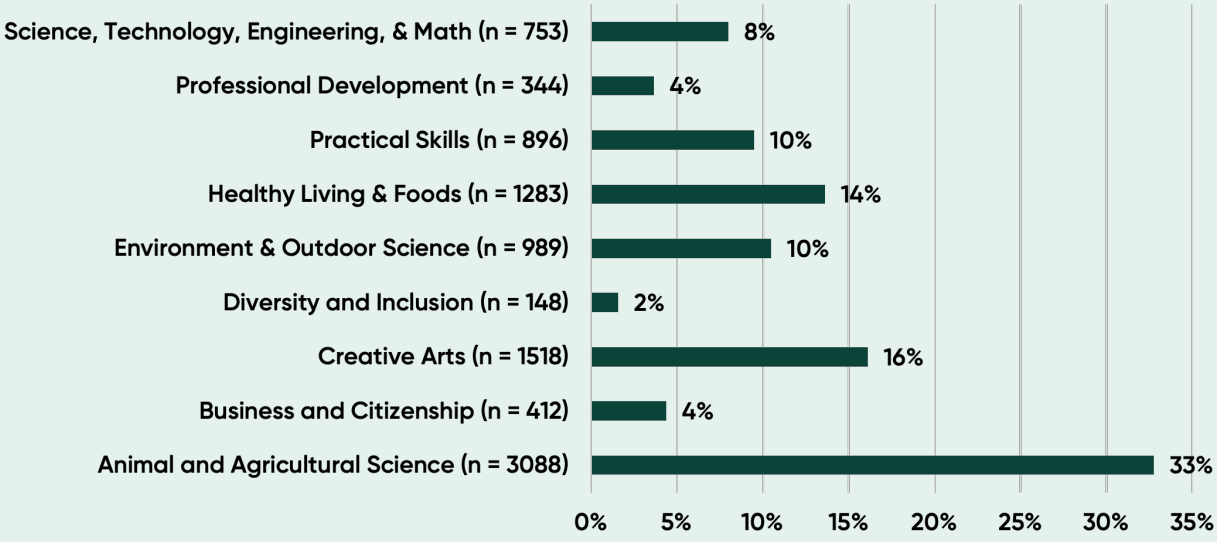


Figure 3. Number of Youth Participants by Project



4-H Involvement: Time, Level and Type

Respondents were asked to rate the level of involvement in 4-H on a 1 (none) to 7 (very high) scale. As shown in Table 4, youth indicated an average rating of 4.62 (*SD* = 1.60) during the school year and 5.37 (*SD* = 1.68) during the summer. The difference between these ratings (Mean difference = -.749, *SE* = .028, 95% CI: -.805, -.694) was statistically significant [*t* (3936) = 26.579, *p* < .001] **indicating higher levels of involvement in the summer**.

Table 4. Level of 4-H Involvement by Time of Year

	Mean	SD
Level of 4-H Involvement During School Year	4.62	1.6
Level of 4-H Involvement During Summer	5.37	1.68

Similarly, respondents reported on their levels of 4-H participation at the county, state, and national level on a 1 (none) to 7 (very high) scale. **The greatest level of participation was at the county level** which was significantly greater than state level participation (Mean Difference = 2.372 *SE* = .031, 95% CI: 2.311, 2.432, *t*(3931) = 77.254, *p* < .001) and national level participation (Mean Difference = 3.515, *SE* = .030, 95% CI: 3.456, 3.575, *t*(3932) = 116.494, *p* < .001). Similarly, reported state level participation was significantly higher than national level participation (Mean Difference = 1.144, *SE* = .024, 95% CI: 1.097, 1.192, *t*(3941) = 47.468, *p* < .001). **Put simply, respondents reported greater rates of involvement in the summer, and reported greatest levels of involvement at the county level**.

Table 5. 4-H Participation by Level

	Mean	SD
Rate of 4-H Participation at County Level	4.87	1.81
Rate of 4-H Participation at State Level	2.50	1.75
Rate of 4-H Participation at National Level	1.36	0.997

To understand the factors that may be related to greater levels of 4-H involvement, a series of bivariate correlations were examined. Because of low numbers of respondents who identified a racial/ethnic group other than white (see Table 3), this variable was recoded for analyses as a continuous variable (0 = White; 1 = Non-White)[1]. Similarly, due to low levels of respondents identifying as a gender other than male or female, respondents who did not identify as either male (0) or female (1) were excluded from this analysis. Furthermore, the strength of relationships between variables were assessed utilizing the guidelines illustrated in Table 6a. In the maximum likelihood framework (e.g., NHST), statistical significance is not important when paired with a weak effect size, especially with larger samples (Cohen et al., 2003). Associations that are negligible (i.e., *r* < .200) should be interpreted as indicating no meaningful relation within the present study and sample. As such, the effect size should be considered rather than the statistical significance when making decisions about how to implement the reported data in their own work (see Table 6a).

Table 6a. Pearson Bivariate Correlation Coefficient Effect Size Guide

Pearson r level	Relationship Strength
> .700	Very Strong Relationship
.400-.699	Strong Relationship
.300-.399	Moderate Relationship
.200-.299	Weak Relationship
.010-.199	No or negligible relationship

Note: Strength of relationship is bidirectional and applies to both negative and positive relations.

[1]The research team recognizes the serious limitations this approach presents regarding interpretation of the study findings (and the lack of homogeneity with a "non-white" racial category).

There were no statistically meaningful relations between gender or participant race and levels of involvement during the school year or during the summer. Similarly, there were no significant relations between respondent reported gender or dummy coded participant race and rates of 4-H participation at the county, state, or national level.

Aa shown in Table 6b, greater total years of 4-H involvement were positively associated with greater levels of 4-H involvement during the school year (r = .197, p < .001) and during the summer (r = .280, p < .001). Similarly, higher levels of total 4-H activities participated in were positively associated with greater levels of 4-H involvement during the school year (r = .211, p < .001), and greater rates of 4-H participation at the state level (r = .280, p < .001).

Higher levels of 4-H involvement during the school year were positively associated with greater rates of participation in 4-H at the state (r = .355, p < .001) and national level (r = .230, p < .001). Similarly, higher levels of 4-H involvement during the summer were positively associated with greater rates of 4-H participation at the state level (r = .296, p < .001), but not meaningfully so at the national level (r = .136, p < .001).

Table 6b.

Pearson Bivariate Correlations Between Demographic Variables and Involvement

Correlations	Mean [SD]	1	2	3	4	5	6	7	8	9	10	11
1. Total Years in 4-H	4.05 [1.49]	--										
2. Total Number of 4-H Activities Participated In	2.38 [1.68]	.268*	--									
3. Bio Sex: 0 = Male; 1 = Female	0.66 [0.47]	-.021 (.191)	.056*	--								
4. Age in Years	15.13 [1.6]	.257*	.111*	.057*	--							
5. Age of Start at 4-H	11.08 [1.88]	-.570*	-.117*	.066*	.648*	--						
6. Dummy Code Participant Race (0 = White; 1 = Non-White)	0.10 [0.30]	-.133*	.015 (.361)	.028 (.088)	-.002 (.881)	.102*	--					
7. Level of 4-H Involvement During School Year	4.62 [1.60]	.197*	.211*	.081*	.094*	-.074*	.019 (.231)	--				
8. Level of 4-H Involvement During Summer	5.37 [1.68]	.291*	.154*	.085*	.082*	-.158*	-.090*	.421*	--			
9. Rate of 4-H Participation at County Level	4.87 [1.81]	.280*	.213*	.068*	.091*	-.142*	-.069*	.464*	.542*	--		
10. Rate of 4-H Participation at State Level	2.5 [1.75]	.197*	.280*	.063*	.144*	-.032 (.048)	.011 (.489)	.355*	.296*	.415*	--	
11. Rate of 4-H Participation at National Level	1.36 [1.00]	.027 (.093)	.094*	.035 (.031)	.128*	.088*	.055*	.230*	.136*	.187*	.507*	--

Note: * indicates p < .001, otherwise all exact p-values provided in parentheses below correlation.

4-H Program Outcomes: Common Measures

Youth responded to selected items from established 4-H Common Measures instruments in the areas of Civic Engagement (4 items), College and Career Readiness (4 items), Healthy Living (4 items), and Science (4 items).

Common Measures Outcomes Descriptive Analysis

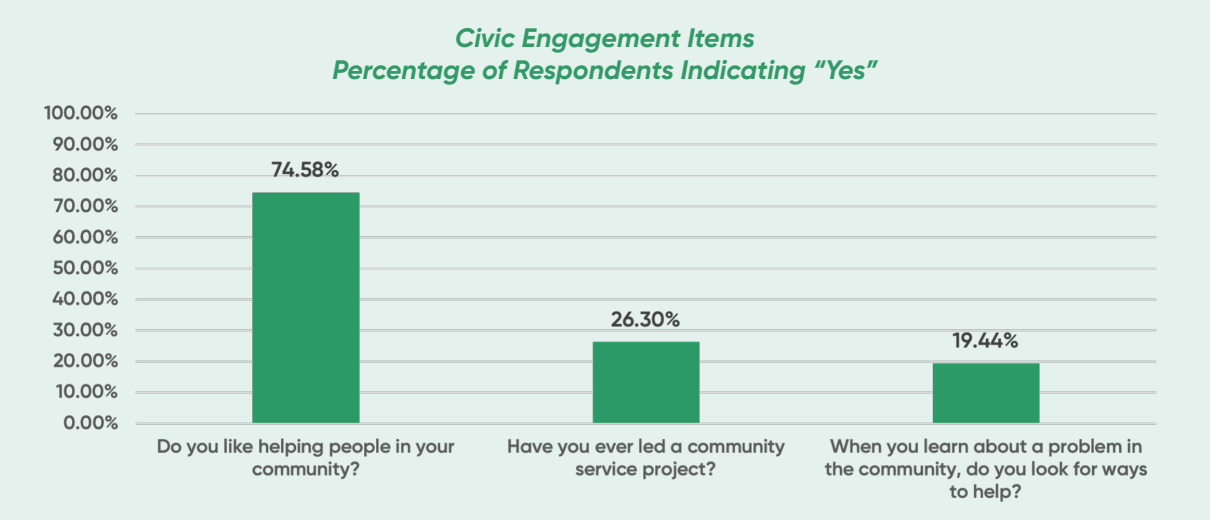
The each of the tables below (7a – 7d) presents the frequencies and percentages of responses for each response category for each item. Each table is followed by a figure (5a -5d) showing the percentage of respondents who rated the items the highest.

CIVIC ENGAGEMENT

Table 7a. Frequencies and Percentages of Responses: Civic Engagement Common Measure Program Outcomes

Civic Engagement	Yes	A Lot	Sort Of	Usually	A Little	Not Really	Not at All	No
Do you like helping people in your community?	2,949 74.58%	--	--	897 22.69	--	97 2.45%	--	11 0.28%
Have you ever led a community service project?	1,039 26.30%	--	--	1,301 32.93%	--	--	--	1,611 40.77%
When you learn about a problem in the community, do you look for ways to help?	767 19.44%	--	--	1,812 45.92%	--	1,250 31.68%	--	117 2.97%
How much has 4-H inspired you to volunteer in your community?	--	1,671 42.27%	--	--	2,011 50.87%	--	271 6.86%	--

Figure 5a. Percentage of Respondents Indicating "Yes": Civic Engagement Common Measure Program Outcomes



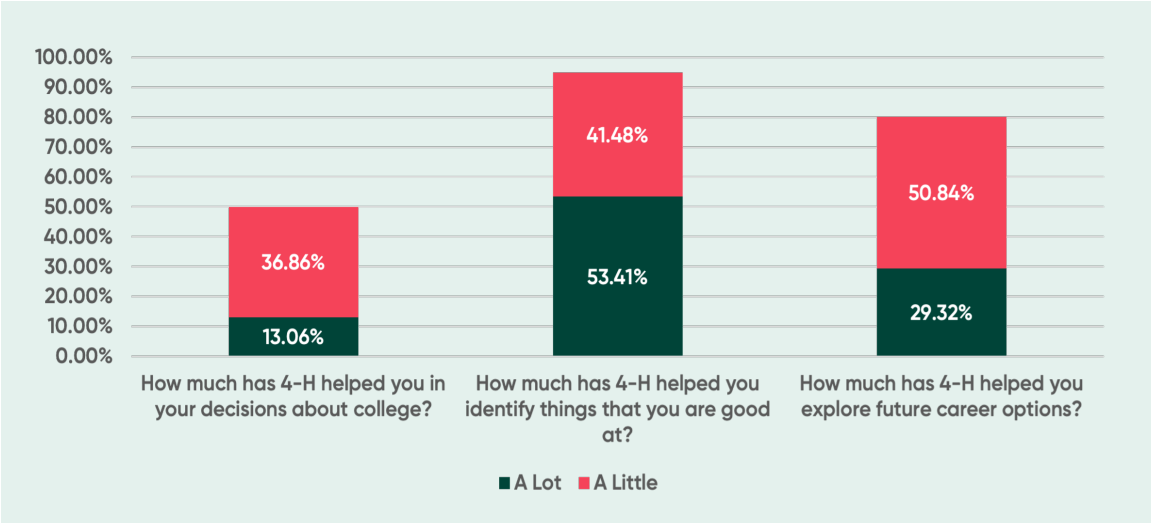
Nearly 90% of the survey participants reported that 4-H had motivated them to engage in community volunteering. As shown in Table 8A, this effect was positively linked to a higher frequency of participants leading community service initiatives (ρ = .329, $p < .001$), a stronger inclination to seek solutions for community issues (ρ = .427, $p < .001$), and a heightened enjoyment in assisting others within their community (ρ = .332, $p < .001$). In simpler terms, **those who were encouraged to volunteer by 4-H were more likely to actively and significantly participate in community service activities**.

COLLEGE AND CAREER READINESS

Table 7b. Frequencies and Percentages of Responses: College and Career Readiness Common Measure Program Outcomes

College and Career Readiness	Yes	A Lot	Sort Of	Usually	A Little	Not Really	Not at All	No
At 4-H, did you learn about colleges that might be a good fit for you?	509 12.88%	--	982 24.85%	--	--	--	--	2,416 62.27%
How much has 4-H helped you in your decisions about college?	--	514 13.06%	--	--	1,451 36.86%	--	1,971 50.08%	--
How much has 4-H helped you identify things that you are good at?	--	2,109 53.41%	--	--	1,638 41.48%	--	202 5.12%	--
How much has 4-H helped you explore future career options?	--	1,157 29.32%	--	--	2,006 50.84%	--	783 19.84%	--

Figure 5b. Percentage of Respondents Indicating "A Lot" of "A Little": College and Career Readiness Common Measure Program Outcomes



The findings showed that the participants' 4-H experiences did not significantly contribute to their understanding of suitable colleges or their college decision-making process. This apparent lack of impact might be attributed to the varied ages and school grades of the participants in the study. To further explore this, we employed an advanced linear regression technique, tailored to account for the categorical nature of the Common Measures items, to examine if there was a correlation between the age of the respondents and their choices in the survey. Interestingly, our analysis revealed a negative correlation between the ages of the respondents and their answers on the items related to college and career readiness. It's important to note that lower scores on these Common Measures items indicate a higher level of agreement with the statements.

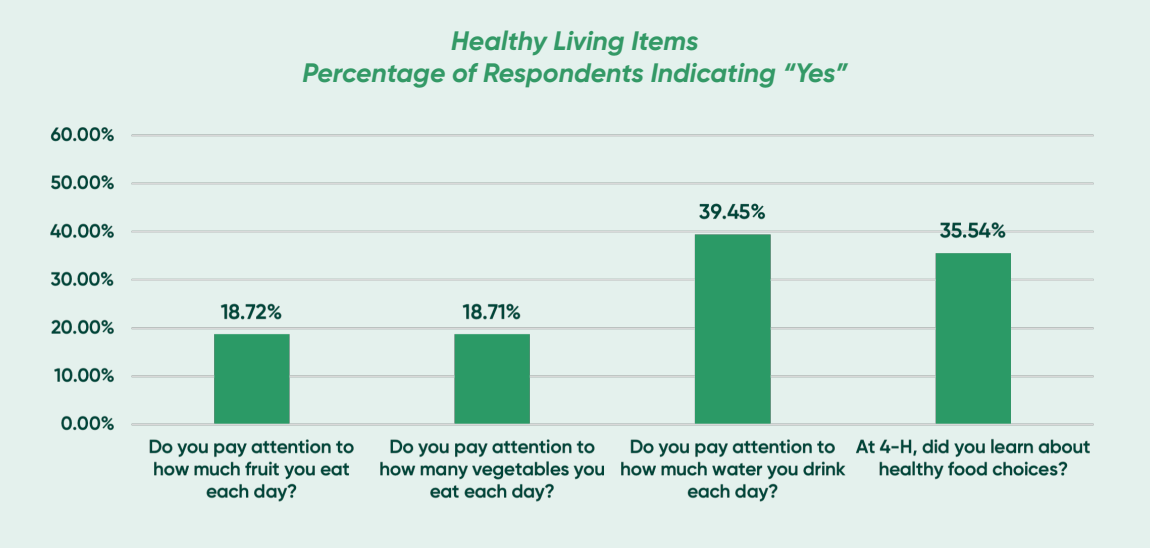
As participants increased in age, they were more likely to report they learned about colleges that may be a good fit at 4-H (β = -.142, *SE* = .007, *p* < .001) and **more likely to report 4-H has helped with decisions about college** (β = -.179, *SE* = .007, *p* < .001). More simply, **4-H was more helpful for older youth in determining their college path.**

HEALTHY LIVING

Table 7c. Frequencies and Percentages of Responses: Healthy Living Common Measure Program Outcomes

Healthy Living	Yes	A Lot	Sort Of	Usually	A Little	Not Really	Not at All	No
Do you pay attention to how much fruit you eat each day?	740 18.72%	--	--	1,086 27.47%	--	1,528 38.64%	--	600 15.17%
Do you pay attention to how many vegetables you eat each day?	739 18.71%	--	--	1,025 25.95%	--	1,568 39.70%	--	618 15.65%
Do you pay attention to how much water you drink each day?	1,559 39.45%	--	--	1,152 29.15%	--	915 23.15%	--	326 8.25%
At 4-H, did you learn about healthy food choices?	1,405 35.54%	--	1,331 33.67%	--	--	--	--	1,217 30.79%

Figure 5c. Percentage of Respondents Indicating "Yes": Healthy Living Common Measure Program Outcomes



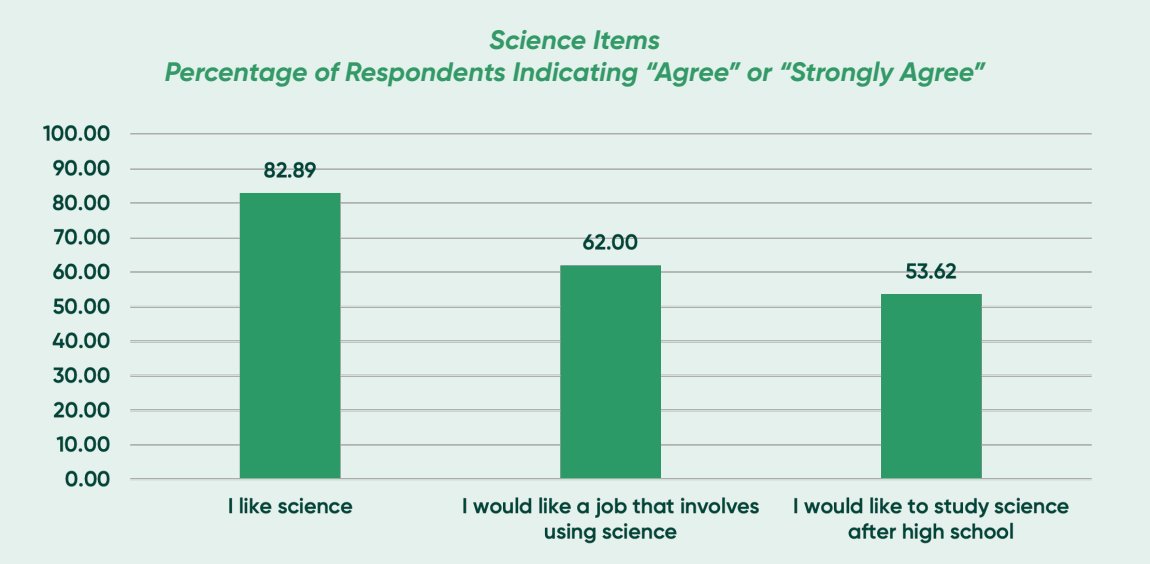
A large proportion of respondents indicated they learned about healthy food choices at 4-H. Learning about these food choices was positively associated with paying attention to eating fruit (ρ = .198, *p* < .001) and paying attention to eating vegetables (ρ = .200, *p* < .001). In summary, **respondents generally indicated their learning about eating healthier was associated with learning about healthy food choices at 4-H.**

SCIENCE

Table 7d. Frequencies and Percentages of Responses: Science Common Measure Program Outcomes

	Strongly Agree	Agree	Disagree	Strongly Disagree
I like science.	835 25.92%	1,835 56.97%	469 14.56%	82 2.55%
I would like a job that involves using science.	577 17.98%	1,413 44.02%	970 30.22%	250 7.79%
I would like to study science after high school.	506 15.78%	1,213 37.84%	1,117 34.84%	370 11.54%

Figure 5d. Percentage of Respondents Indicating "Agree" or "Strongly Agree" for Science Program Outcomes



Around two-thirds of the respondents shared that they engaged in discussions and expanded their knowledge about science, as well as applied it to solve daily challenges, through their participation in 4-H. This emphasis on science was positively correlated with the decision to pursue a career in science after completing high school ($\rho = .240$, $p < .001$). **A significant portion of those involved in 4-H reported that learning, utilizing, and appreciating science were key aspects of their experience with the program.**

Common Measures Outcomes Between Variables Correlational Analysis

As noted in Hawley et al. (2016), the Common Measures items utilized differing response categories both within and across the four Common Measures content areas (e.g., Yes, Usually, Not Really, No versus A lot, A little, Not at all). Because of the expected challenges related to the differing response categories, a non-parametric (i.e., non-maximum likelihood) approach was selected for analyses of relations between and across study variables where appropriate.

Because of the categorical nature of the Common Measure items Spearman's rank-order correlations (ρ) (also known as Spearman's rho) were employed to assess the relations between the four selected Common Measure outcome areas, number of 4-H activities participated in, participant gender identity and participant racial identity (Cohen et al., 2003). The strength of relationships between variables were assessed utilizing the guidelines presented in Table 8a. Importantly, and as noted earlier, statistical significance is not important when paired with a weak effect size, especially with larger samples (Cohen et al., 2003).

Table 8a. Spearman Rank-Order Correlation Coefficient Effect Size Guide

Pearson r level	Relationship Strength
> .700	Very Strong Relationship
.400-.699	Strong Relationship
.300-.399	Moderate Relationship
.200-.299	Weak Relationship
.010-.199	No or negligible relationship

Note: Strength of relationship is bidirectional and applies to both negative and positive relations.

Additionally, three items from the science content area were measured on a 1 (Strongly Agree) to 4 (Strongly Disagree) scale. To enhance parsimony in reporting, a composite score of these three items was created (by summing the items and dividing by 3) which demonstrated acceptable levels of internal consistency (α = .885).

The strength of Spearman's correlation coefficients (ρ) across the 17 variables assessed was generally weak or negligible when these variables were correlated with those outside their specific construct area, as shown in Table 8A. For instance, items related to healthy living showed strong correlations with other items within the same category but had weak to negligible relationships with items from different constructs. Nevertheless, the data did illustrate a few significant relationships worth noting.

There was a positive relationship between "how much 4-H inspired your volunteering" (i.e., civic engagement) and identifying colleges that would be a good fit (ρ = .327, $p < .001$), helping in decisions about college (ρ = .369, $p < .001$), identifying things they are good at (ρ = .399, $p < .001$), exploring future career options (ρ = .361, $p < .001$), and talking about science to solve everyday problems (ρ = .315, $p < .001$). **As respondents reported higher levels of volunteer inspiration, they also reported higher levels of college and career readiness and higher levels of using science to solve everyday problems**.

The college and career readiness variables were also positively associated with talking about science to solve everyday problems at a moderate level. **More specifically, talking about science was positively associated with learning about colleges that might be a good fit** (ρ = .315, $p < .001$), 4-H helping with decisions about college (ρ = .328, $p < .001$), identifying things you are good at (ρ = .283 $p < .001$), and exploring future career options (ρ = .271, $p < .001$).

Crucially, there were no statistically significant correlations (beyond weak or negligible levels, as shown in Table 8a) between the participants' self-reported gender (coded as a dummy variable), racial identity (also coded as a dummy variable), the number of 4-H projects they participated in, and any of the standard 4-H evaluation measures. In other words, a respondent who identified as a non-white individual had an equal likelihood of reporting a "high" score in science application as a respondent who identified as white.

Table 8b.

Spearman's Rank Order Correlations of Common Measures Data and Respondent Demographics

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1. Do you like helping people in your community?	--																	Civic Engagement
2. Have you ever led a community service project?	.237 *	--																
3. When you learn about a problem in the community, do you look for ways to help?	.336 *	.384 *	--															
4. How much has 4-H inspired you to volunteer in your community?	.332 *	.329 *	.427 *	--														
5. At 4-H, did you learn about colleges that might be a good fit for you?	.161 *	.214 *	.261 *	.327 *	--													College & Career Readiness
6. How much has 4-H helped you in your decisions about college?	.166 *	.219 *	.243 *	.369 *	.620 *	--												
7. How much has 4-H helped you identify things that you are good at?	.180 *	.207 *	.232 *	.399 *	.369 *	.468 *	--											
8. How much has 4-H helped you explore future career options?	.149 *	.201 *	.216 *	.361 *	.469 *	.556 *	.538 *	--										
9. Do you pay attention to how much fruit you eat each day?	.143 *	.135 *	.218 *	.176 *	.149 *	.133 *	.104 *	.125 *	--									Healthy Choices
10. Do you pay attention to how many vegetables you eat each day?	.139 *	.144 *	.230 *	.180 *	.144 *	.123 *	.099 *	.119 *	.884 *	--								
11. Do you pay attention to how much water you drink each day?	.141 *	.136 *	.215 *	.168 *	.125 *	.133 *	.129 *	.125 *	.493 *	.491 *	--							
12. At 4-H, did you learn about healthy food choices?	.168 *	.153 *	.218 *	.332 *	.293 *	.259 *	.290 *	.310 *	.198 *	.200 *	.178 *	--						
13. Did you talk about how science can be used to help solve everyday problems?	.160 *	.178 *	.242 *	.315 *	.328 *	.283 *	.271 *	.337 *	.137 *	.144 *	.119 *	.449 *	--					Science
14. Composite of Science Items	.115 *	.103 *	.146 *	.102 *	.128 *	.165 *	.089 *	.105 *	.132 *	.136 *	.057 *	.047 (.003)	.240 *	--				
15. Total number of 4-H Activities Participated In	-.093 *	-.278 *	-.142 *	-.206 *	-.115 *	-.140 *	-.162 *	-.144 *	-.066 *	-.083 *	-.039 (.014)	-.271 *	-.225 *	-.130 *	--			
16. Respondent Report Gender** 0 = Male; 1 = Female	-.115 *	-.111 *	-.133 *	-.118 *	-.073 *	-.111 *	-.097 *	-.101 *	-.152 *	-.145 *	-.121 *	-.045 (.005)	.046 (.004)	.008 (.630)	-.053 *	--		
17. Dummy Coded Participant** Race (0 = White; 1 = Non-White)	-.013 (.430)	.009 (.560)	-.028 (.086)	-.069 *	-.049 (.002)	-.038 (.018)	-.012 (.475)	-.023 (.156)	-.044 (.006)	-.041 (.012)	-.058 *	-.035 (.028)	-.046 (.004)	-.040 (.014)	.001 (.943)	.028 (.088)	--	

Note: Exact p-values provided in parentheses below correlation if $p > .001$, other * indicates $p < .001$. **indicates the variable is dummy coded.

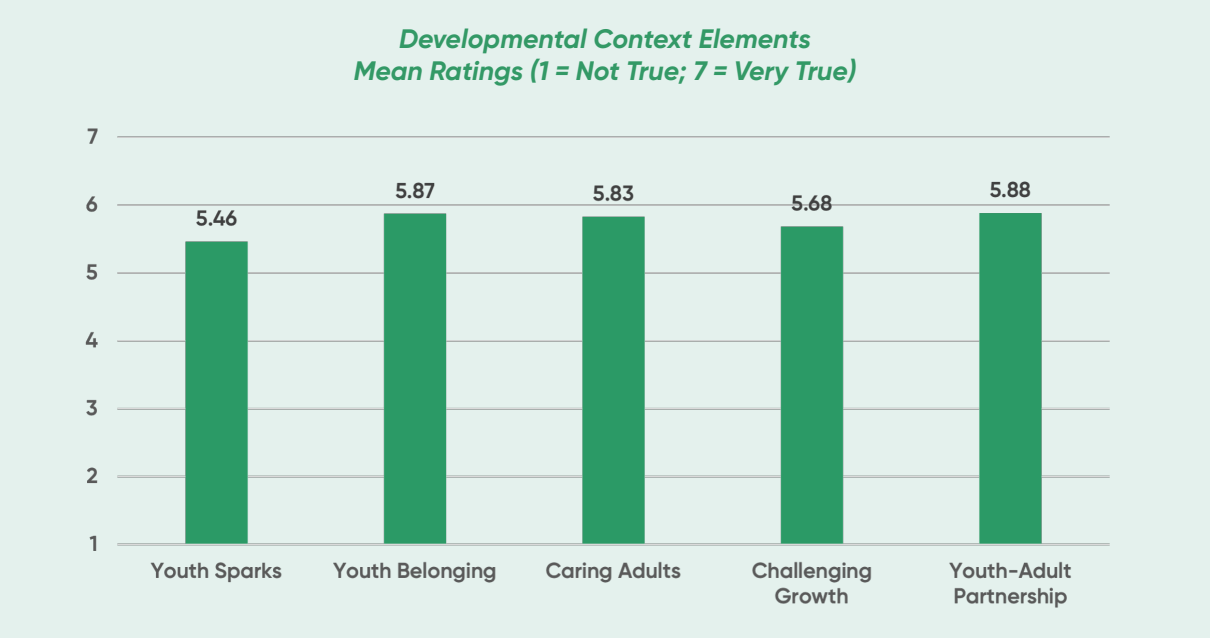
Impact of 4-H Participation on Positive Youth Development

DEVELOPMENTAL CONTEXT

The 4-H Thriving Model shows how 4-H programs can provide a developmental context that leads to positive youth development outcomes. The developmental context is based on a combination of youth having the opportunity to explore their spark and experience a sense of belonging along with building positive developmental relationships with adults and their peers in 4-H. In the 4-H Thriving Model developmental relationships are marked by adults who care, challenge growth, and share power through youth-adult partnerships.

Youth respondents reported high levels across each measure of developmental context (mean scores between 5.455 and 5.877), indicating that on average youth are experiencing high-quality positive **youth development program settings in 4-H**.

Figure 9. Mean Ratings of Developmental Context Elements



In brief, 4-H as a developmental context is a combination of youth feeling like they belong in the program "space," they have supportive relationships with their peers and adults at 4-H, and the 4-H program is a space where they are consistently growing. **Respondents who scored higher on developmental context tended to also report greater levels of inspiration to serve their communities (ρ = .404, *p* < .001), better identification of their strengths (ρ = .457, *p* < .001), and better understanding of future career options (ρ = .379, *p* < .001).** 4-H as a developmental context was also positively associated with greater levels of learning about healthy food choices (ρ = .292, *p* < .001) and the utilization of science to solve everyday problems (ρ = .290, *p* < .001).

When 4-H programs facilitate youth feeling like they belong in the program "space," they have supportive relationships with their peers and adults at 4-H, and the 4-H program is a space where they are consistently growing, they also tend to score "better" across the four dimensions of Common Measures items.

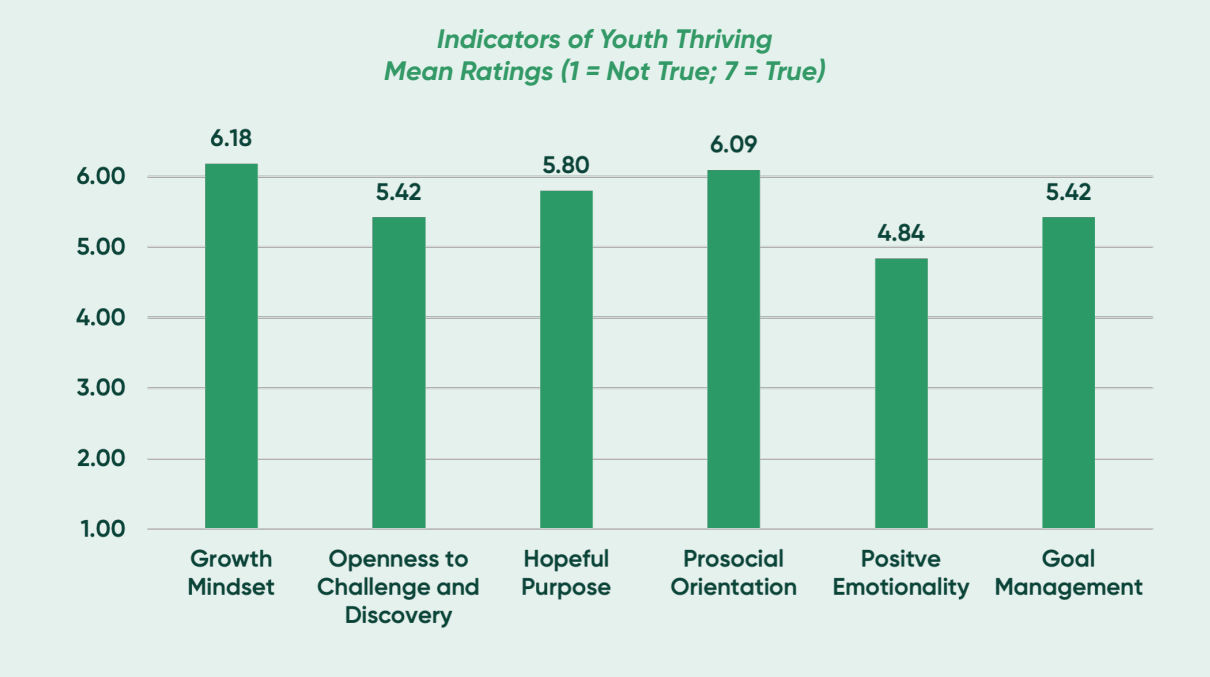
YOUTH THRIVING

The seven indicators of youth thriving articulated in the 4-H Thriving Model represent the research-based social, emotional, cognitive, and behavioral skills, along with the positive habits of mind, that mark movement on a positive trajectory toward adulthood. **These seven indicators are**:

1. openness to challenge and discovery;
2. growth mindset;
3. prosocial awareness;
4. hopeful purpose;
5. transcendent awareness;
6. positive emotionality; and
7. intentional self-regulation through goal setting and management.

Youth respondents reported high levels across each measure of thriving (mean scores between 4.844 and 6.177), indicating that on average **youth are experiencing high-quality, positive youth development program settings in 4-H**. It is worth noting that the score for positive emotionality is lower than the others. **At 4.9 this score is still considered in the high range, but given its variation from the other indicators, could reflect the ongoing mental health concerns of young people**.

Figure 10. Mean Ratings of Thriving Indicators



When a youth thrives because of their 4-H experience, they are more open to challenges, more curious, higher in emotional intelligence, are focused on self-improvement, and goal oriented. Youth who reported higher levels of thriving also reported greater levels of civic engagement, more specifically youth reported greater levels of helping in their community ($\rho = .310$, $p < .001$), looking for ways to help with problems in their community ($\rho = .321$, $p < .001$), and greater levels of inspiration to volunteer in their community ($\rho = .332$, $p < .001$). Additionally, greater levels of reported youth thriving were associated with youth reporting that 4-H helped them identify things they are good at ($\rho = .299$, $p < .001$).

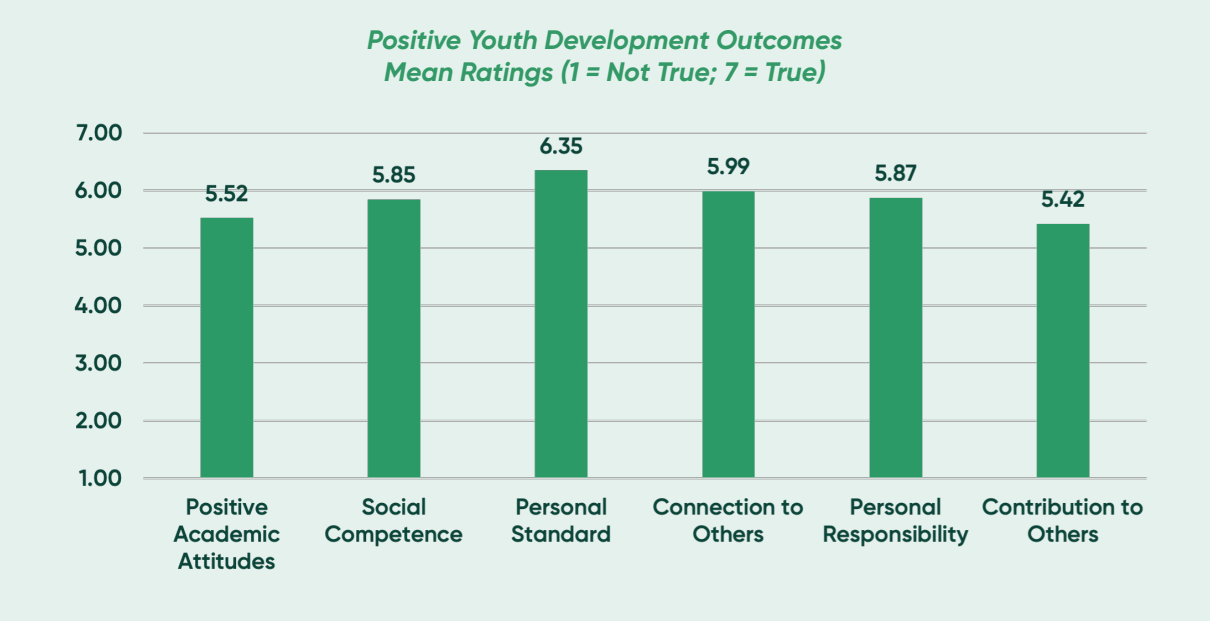
In summary, when a youth experiences the dimensions of youth thriving centered programming at 4-H (e.g., openness to challenges, greater curiosity, practiced in emotional intelligence, a focus of self-improvement, and goal orientation), they also report greater levels of civic engagement, college and career readiness, healthy living, and enjoyment of science.

POSITIVE YOUTH DEVELOPMENTAL OUTCOMES

Youth participation in 4-H programs that provide a high-quality developmental setting and opportunities to build thriving indicators result in key positive youth development outcomes. These outcomes include positive attitudes toward academic performance, social competence, high personal standards, a connection to others (peers, adults and their community), a sense of personal responsibility, and dedication to contributing to others and community.

Youth respondents reported high levels of attainment for all PYD outcomes (mean scores between 5.419 and 6.350), indicating that on average **youth are achieving important developmental outcomes through participation in 4-H**.

Figure 11. Mean Ratings of Developmental Outcomes



When youth thrive because of their 4-H experiences, they are more open to challenges, more curious, higher in emotional intelligence, focused on self-improvement, and are goal oriented. Youth who reported higher levels of thriving also reported greater levels of civic engagement, more specifically youth reported greater levels of helping in their community ($\rho = .310$, $p < .001$), looking for ways to help with problems in their community ($\rho = .321$, $p < .001$), and greater levels of inspiration to volunteer in their community ($\rho = .332$, $p < .001$). Additionally, greater levels of reported youth thriving were associated with youth reporting that 4-H helped them identify things they are good at ($\rho = .299$, $p < .001$).

In summary, when a youth experiences the dimensions of youth thriving centered programming at 4-H (e.g., openness to challenges, greater curiosity, practiced in emotional intelligence, a focus of self-improvement, and goal orientation), they also report greater levels of civic engagement, college and career readiness, healthy living, and enjoyment of science.

Relationship Between Common Measures and Youth Thriving

The potential associations between Common Measures sample items and the three dimensions of the youth thriving model were also examined. As illustrated in Table 12 and described below, there were a variety of weak, moderate, and strong relations present.

Table 12. Spearman's Rank Order Correlations of Youth Thriving Dimensions, Common Measures Data and Respondent Demographics

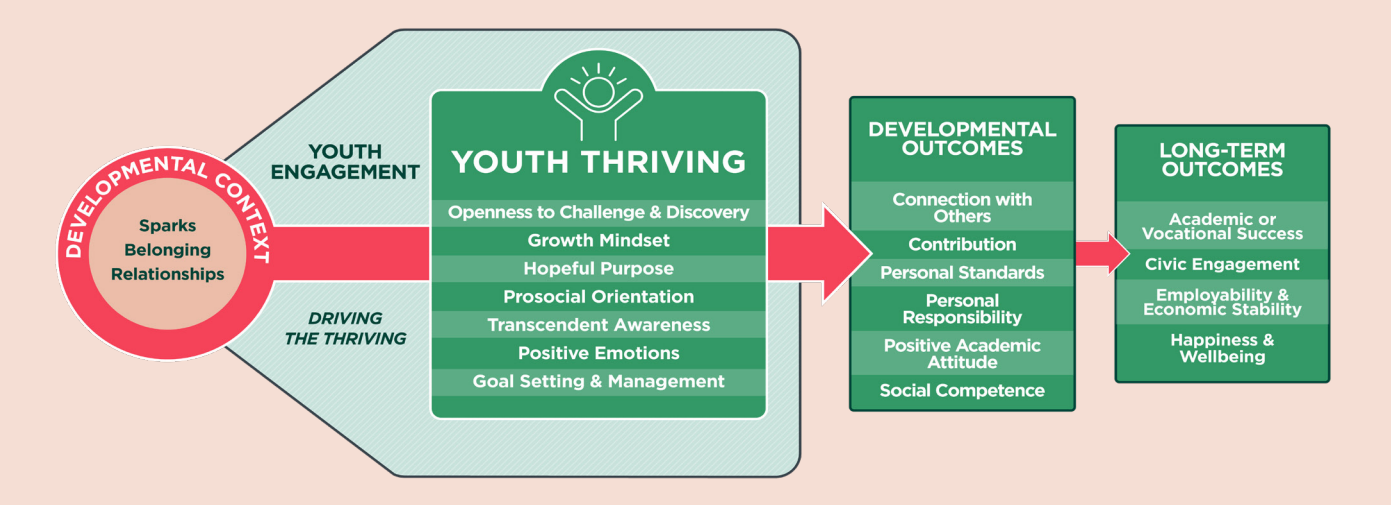
Common Measures Domain	Common Measures Item	Developmental Context	Developmental Outcomes	Youth Thriving
Civic Engagement	1. Do you like helping people in your community?	.241	.383	.310
	2. Have you ever led a community service project?	.167	.284	.207
	3. When you learn about a problem in the community, do you look for ways to help?	.219	.376	.321
	4. How much has 4-H inspired you to volunteer in your community?	.404	.404	.332
College and Career Readiness	5. At 4-H, did you learn about colleges that might be a good fit for you?	.288	.228	.223
	6. How much has 4-H helped you in your decisions about college?	.338	.266	.251
	7. How much has 4-H helped you identify things that you are good at?	.457	.304	.299
	8. How much has 4-H helped you explore future career options?	.379	.247	.257
Healthy Living	9. Do you pay attention to how much fruit you eat each day?	.092	.244	.222
	10. Do you pay attention to how many vegetables you eat each day?	.088	.259	.232
	11. Do you pay attention to how much water you drink each day?	.129	.266	.255
	12. At 4-H, did you learn about healthy food choices?	.292	.222	.205
Science	13. At 4-H, did you talk about how science can be used to help solve everyday problems?	.290	.223	.227
	14. Composite of Science Items (See Table 7)	.116	.225	.207
n/a	15. Total number of 4-H Activities Participated In	-.097	-.117	-.074
n/a	16. Respondent-Reported Gender** (0 = Male; 1 = Female)	-.046	-.169	-.119
n/a	17. Dummy Coded Participant** Race (0 = White; 1 = Non-White)	.026		

*Note: Exact p-values provided in parentheses below correlation if p > .001, other * indicates p < .001. **indicates the variable is dummy coded; The Thriving Items were reverse coded (only) for this analysis to be on the same "scaling" as the Common Measures' items.*

Replication of the 4-H Thriving Model

The 4-H Thriving Model (Arnold, 2018) advanced the measurement of the impact of 4-H by identifying the processes through which 4-H contributes to the positive development of youth. As shown in Figure 6, The 4-H Thriving Model outlines the connection between participation in high-quality 4-H programs (the developmental context), the effect of that participation on youth thriving and how thriving youth, in turn, achieve key developmental outcomes, setting them up to achieve longer-term outcomes in adulthood. The model has been successfully tested and established as the model that shows how 4-H contributes to the positive development of youth (Arnold & Gagnon, 2019).

Figure 6. The 4-H Thriving Model



The 2022 National Index Study included measurement of the 4-H Thriving Model components consisting of:

- **Developmental Context Items** – that measure the quality of the 4-H program setting, and include the elements of youth sparks, belonging, and developmental relationships. Developmental relationships are measured based on youth experiences with adults that express care, challenge growth, and share power through positive youth-adult partnerships.
- **Indicators of Youth Thriving Items** – that measure the research-based social, emotional, cognitive, and behavioral skills, along with the positive habits of mind, that indicate movement on a positive trajectory toward adulthood. These seven indicators are:
  1. openness to challenge and discovery;
  2. growth mindset;
  3. prosocial awareness;
  4. hopeful purpose;
  5. transcendent awareness;
  6. positive emotionality; and
  7. intentional self-regulation through goal setting and management.

- **Positive Youth Development Outcomes Items** – that measure positive attitudes toward academic performance, social competence, high personal standards, a connection to others (peers, adults, and their community), a sense of personal responsibility and dedication to contributing to others and community.

Each item is measured on a 1-7 scale with "1" indicating the item was not true and "7" indicating the item was very true.

Using this measurement provided the opportunity to replicate the model's structure on a larger, multi-state data set, and to explore the potential associations between 4-H program outcomes and positive youth development for the first time. Data from the National Index Study allowed us to test the model's structure and psychometric properties on a larger, more representative sample and see if the structural model and mediational processes found by Arnold and Gagnon (2019) occur in a larger, more representative sample. Using this measurement provided the opportunity to replicate the model's structure on a larger, multi-state data set, and to explore the potential associations between 4-H program outcomes and positive youth development for the first time. Data from the National Index Study allowed us to test the model's structure and psychometric properties on a larger, more representative sample and see if the structural model and mediational processes found by Arnold and Gagnon (2019) and Gagnon et al. (2022) occur in a larger, more representative sample. The model was replicated in the 2022 National 4-H Index Study, and now again in the 2023 study, leading to continued confidence in the model's structure and use as a theory of change for 4-H Positive Youth Development.

Planned Missing Data Design

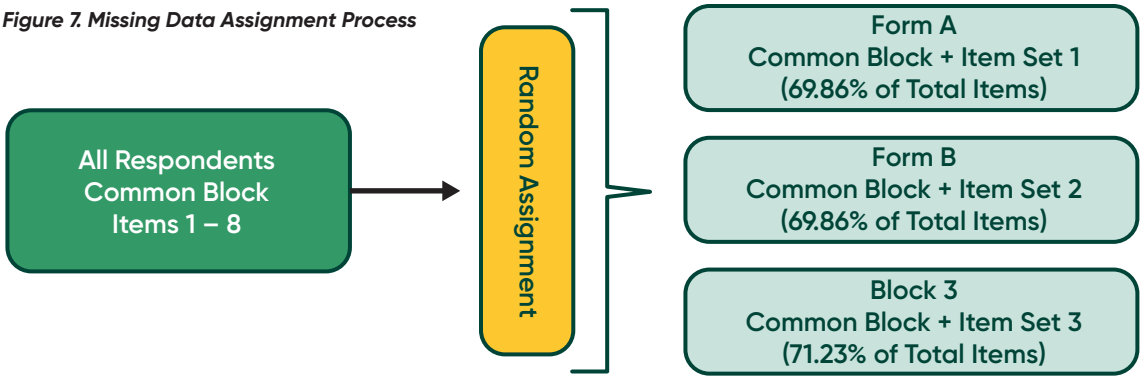
The thriving model (and scale) captures the intricate dynamics of developmental contexts, outcomes, and youth thriving, resulting in a comprehensive scale comprising 73 items. To mitigate respondent fatigue due to the scale's length, a planned missing data approach was adopted. Essentially, this method involves randomly assigning respondents to a version of the survey with fewer questions. As depicted in Table 9 and Figure 7, all respondents answer a core set of questions before being randomly placed into one of three different conditions. This random assignment ensures that any missing data occurs completely at random (MCAR), meaning it's not linked to any specific characteristic of the participant. The MCAR nature of the data permits the use of advanced missing data handling methods. Specifically, in this planned missing data context, the gaps in data are filled using techniques like multiple imputation (MI), expectation maximization (EM), or, as in this study, Full Information Maximum Likelihood (FIML). FIML utilizes all available data points to estimate a respondent's missing values. The specific items of the three different conditions related to the missing data are detailed in Table 11.measured participant characteristic. Further, the MCAR conditioning of the data allows for contemporary missing data management techniques. Specifically, in a planned missing data design, the missing data points are then estimated and/or imputed utilizing a technique such as multiple imputation (MI), expectation maximization (EM), or in the case of the present study, Full Information Maximum Likelihood (FIML). With a FIML technique, all available points of data are employed to estimate a respondent's missing values (thus the "Full" in Full Information Maximum Likelihood). The items associated with the three specific missing data conditions are presented in Table 9.

Table 9. Planned Missingness Design

	Common Items (8-Items) + (Item Set 1, 2 or 3)	Item Set 1 (43-Items)*	Item Set 2 (43-Items)*	Item Set 3 (44-Items)*
Form A	X	X	X	•
Form B	X	•	X	X
Form C	X	X	•	X

Note: All respondents received common items; X indicates respondent received item; • indicates planned missing data; * is total number of items in set excluding common items.

Figure 7. Missing Data Assignment Process



Data Diagnostics and Analytic Plan

The study's framework, which examined the measurement properties of the scale and tested hypotheses, required the assumption of multivariate normality. To check for this, we used the MissMech package (version 1.0.2; Jamshidian et al., 2014) to screen the data for multivariate normality. The analysis revealed that the data were not multivariate normal (indicated by Hawkins's test $p < .001$ and the Anderson Darling k-sample test, $p = .035$). Therefore, we employed robust estimation techniques to compensate for this deviation from normality, crucial for maximum likelihood (ML) analyses. Specifically, we used maximum likelihood estimation with robust standard errors (MLR) to mitigate the risk of Type 1 error and/or misinterpretation of model fit and parameter estimates (Du & Bentler, 2022). We also checked for multivariate outliers using a combination of Mahalanobis distance and the chi-square distribution, which identified 180 respondents as significantly non-normal ($p < .001$). Consequently, two models were generated, both for measurement and hypothesis testing, one including and one excluding these outliers. A comparison of model fits between these two datasets revealed no statistically significant differences, leading to the decision to retain the outliers in the final data set.

Confirmatory factor analysis (CFA) and structural equation modeling (SEM) were used to test the measurement properties and hypotheses, employing the lavaan package (version 0.6-16; Rosseel, 2012) and the semTools package (version 0.5-6, Jorgensen et al., 2022) in R (version 4.3.1). The model fit for both CFA and SEM was evaluated using a robust form of the Root Mean Squared Error of Approximation (RMSEA), with values nearer to zero indicating a more accurate model fit (e.g., RMSEA < .070). The standardized root mean square residual (SRMR) was also used to measure the discrepancy between observed and hypothesized/predicted correlations, where values closer to zero suggest a better fit (e.g., SRMR < .100) (Kline, 2023). Additionally, the Tucker Lewis Index (TLI) and Comparative Fit Index (CFI) were applied to gauge how much the proposed model improves over a baseline model, with scores approaching one being more desirable (e.g., TLI > .900) (Loehlin & Beaujean, 2017). The evaluation of model fit was not based on rigid cutoff scores (e.g., RMSEA = .071 being unacceptable vs. RMSEA = .069 being acceptable); instead, it was contextualized considering the complexity of the model and its performance in previous studies (Chen et al., 2008).

The measurement model was scrutinized for both convergent and discriminant validity, in addition to its fit. For convergent validity, factor loadings were evaluated to determine how well each theorized factor represents its specific items. Here, values closer to one indicate a stronger influence of the factor on the item, whereas lower values ($\lambda < .400$) might suggest a need for redefining the item in the model. Convergent validity was further assessed using McDonald's omega (ω), a measure of internal consistency superior to Cronbach's Alpha (α), as it avoids the often unmet assumption of tau equivalence that α relies on (Hayes & Coutts, 2020). Moreover, the Average Variance Extracted (AVE) was checked to ensure that the factors explained more variance than they did error (i.e., AVE > .500). For discriminant validity, it was important to confirm that the scales represented distinct constructs. This was done by analyzing the correlations between factors; lower correlations suggest that each factor accounts for unique variance (i.e., $r < .700$). Additionally, the square root of the AVE was compared to these correlations to verify that each variable accounted for more unique variance than it shared with other factors in the model ($\sqrt{AVE} > r$).

4-H Thriving Model Confirmatory Factor Analyses

Following the methodology outlined in Arnold and Gagnon (2019), the 73-item, 22-factor scale's psychometric properties were evaluated using confirmatory factor analysis and supplementary statistics. The analysis yielded an acceptable model fit, as indicated by the results: $\chi^2(2533) = 10{,}490.467$, $p < .001$, CFI = .921, TLI = .918, RMSEA = .028 (90% CI from .028 to .029), and SRMR = .052. Table 11 shows that each of the 22 factors had satisfactory internal consistency levels, with ω values ranging from .694 to .947. Additionally, all items/factors exhibited acceptable item loadings, varying from .575 to .962. The AVE values for the three main factors all surpassed .500 (Developmental Context AVE = .684; Developmental Outcomes AVE = .622; Youth Thriving AVE = .631), suggesting the factors account for more variance than error.

Discriminant validity was assessed by comparing the √AVE levels with the correlations between the three primary factors. Table 10 reveals somewhat mixed results regarding the discriminant validity of the three-factor model. This is particularly evident in the high correlation between developmental outcomes and youth thriving ($r = .923$, $p < .001$), although the √AVE for developmental outcomes (√AVE = .788) and youth thriving (√AVE = .794) was greater than their inter-correlation. However, considering previous findings of a strong correlation between these factors (Arnold and Gagnon, 2019) and the overall acceptable psychometric qualities of the measures, the hypothesized model was maintained. Consequently, the proposed mediational model was further explored using structural equation modeling (SEM).

Table 10. Tests of Discriminant Validity

			Between Factor Correlations		
	AVE	√AVE	F1	F2	F3
F1. Developmental Context	0.684	0.827	--		
F2. Developmental Outcomes	0.622	0.788	.572 [.686]	--	
F3. Youth Thriving	0.631	0.794	.541 [.506]	.923 [.781]	--

Note: AVE is Average Variance Extracted; √AVE is the square root of AVE; Correlations between F1, F2, and F3 all p < .001; [Square Bracketed Values] are correlation values from Arnold and Gagnon (2019).

Table 11. Descriptive Data from Confirmatory Factor Analyses

Factor/Item	M (SD)	ω	λ
Developmental Context (Blended 2nd and 3rd Order Factor)**		.749	
Developmental Relationships*			.682
Youth Belonging			.938
Youth Sparks			.842
Developmental Relationships*		.927	
Caring Adults			.964
Challenging Growth			.943
Youth-Adult Partnerships			.907
Youth Sparks		.901	
4-H gives me the opportunity to explore something I really care about. [ABC]	5.61 (1.47)		.795
I am passionate about the things I do in 4-H. [ABC]	5.73 (1.43)		.882
I want to learn all I can about the topic of my 4-H program. [ABC]	5.36 (1.54)		.821
4-H is an important part of who I am. [ABC]	5.12 (1.75)		.824
Youth Belonging		.922	
I feel welcome in this 4-H program. [ABC]	5.93 (1.48)		.874
I feel safe in this 4-H program. [ABC]	6.30 (1.19)		.773
I feel supported by other kids in 4-H. [ABC]	5.58 (1.56)		.862
I feel like I matter in 4-H. [ABC]	5.64 (1.60)		.917
Caring Adults		.945	
At 4-H adults pay attention to me. [BC]	5.83 (1.41)		.893
At 4-H adults like me. [AC]	5.94 (1.36)		.861
At 4-H adults invest time in me. [AB]	5.80 (1.48)		.919
At 4-H adults show an interest in me. [BC]	5.76 (1.47)		.956
Challenging Growth		.913	
At 4-H adults help me see future possibilities for myself. [AC]	5.35 (1.70)		.877
At 4-H adults expect me to do something positive with my future. [AB]	6.04 (1.39)		.840
At 4-H adults stretch me and push me in new ways. [BC]	5.50 (1.65)		.891
At 4-H adults hold me accountable. [AC]	5.81 (1.45)		.787
Youth-Adult Partnerships		.947	
At 4-H adults listen to my ideas. [AB]	5.61 (1.60)		.875
At 4-H adults treat me fairly. [BC]	6.01 (1.39)		.899
At 4-H adults take me seriously. [AC]	5.88 (1.42)		.932
At 4-H adults respect me. [AB]	6.05 (1.35)		.916

*Note: * indicates 2nd Order Factor; ** indicates 3rd Order Factor; λ indicates factor loading; ω indicates Composite Reliability (i.e., Omega); Within 2nd and 3rd order factors, 1st and 2nd order factors act as "items." A, B and C superscript indicate missing data condition.*

Table 11. Descriptive Data from Confirmatory Factor Analyses (Continued)

Factor/Item	M (SD)	ω	λ
Developmental Outcomes**		.860	
Positive Academic Attitudes			.674
Social Competence			.806
Personal Standards			.873
Connection to Others			.802
Personal Responsibility			.793
Contribution to Others			.772
Positive Academic Attitudes		.891	
I think the things I learn in school are useful. [BC]	5.21 (1.42)		.671
Being a student is one of the most important parts of who I am. [AC]	4.93 (1.67)		.729
I want to learn as much as I can at school. [AB]	5.45 (1.46)		.832
I think it is important to earn good grades. [BC]	6.08 (1.29)		.800
I think a lot about how to do well in school. [AC]	5.52 (1.58)		.834
School is very important for later success. [AB]	5.91 (1.39)		.767
Social Competence		.841	
I get along well with people who are different than me. [BC]	5.79 (1.24)		.703
I listen to the opinions of others. [AC]	5.71 (1.20)		.768
I control my anger when I have a disagreement with someone. [AB]	5.54 (1.29)		.695
I follow the rules when I am in a public setting. [BC]	6.40 (.886)		.652
I respect the views of others, even if I disagree. [AC]	5.80 (1.18)		.775
Personal Standards		.869	
It is important for me to do the right thing. [AB]	6.51 (.823)		.754
It is important for me to be a role model for others. [BC]	6.09 (1.20)		.803
It is important for me to do my best. [AC]	6.44 (.90)		.811
It is important that others can count on me. [AB]	6.38 (.94)		.837
Connection to Others		.819	
I have people in my life to whom I look up and admire. [BC]	6.41 (1.02)		.614
I think it is important to be involved with other people. [AC]	6.00 (1.19)		.800
Having friends is important to me. [AB]	6.21 (1.17)		.652
I feel connected to my friends. [BC]	5.97 (1.24)		.642
I feel connected to others in my community. [AC]	5.29 (1.44)		.777
Personal Responsibility		.879	
I take responsibility for my actions. [AB]	6.07 (1.03)		.774
I can be counted on to follow through on things I say I will do. [BC]	6.02 (1.08)		.863
I am a responsible person. [AC]	6.11 (1.05)		.835
I do the things I promise to do without being reminded. [AB]	5.33 (1.31)		.780

*Note: * indicates 2nd Order Factor; ** indicates 3rd Order Factor; λ indicates factor loading; ω indicates Composite Reliability (i.e., Omega); Within 2nd and 3rd order factors, 1st and 2nd order factors act as "items." A, B and C superscript indicate missing data condition.*

Table 11. Descriptive Data from Confirmatory Factor Analyses (Continued)

Factor/Item	M (SD)	ω	λ
Contribution to Others		.904	
I volunteer in my community. [BC]	5.19 (1.67)		.770
Giving back to my community is important to me. [AC]	5.35 (1.43)		.916
It is important for me to contribute my time to help others. [AB]	5.52 (1.36)		.908
I have things that I can contribute for the well-being of others. [BC]	5.57 (1.32)		.801
Youth Thriving**		.806	
Hopeful Purpose			.795
Prosocial Orientation			.824
Positive Emotionality			.693
Challenge & Discovery			.773
Goal Management			.875
Challenge & Discovery*		.694	
Growth Mindset			.793
Openness to Challenge & Discovery			.767
Growth Mindset		.870	
No matter how intelligent I am, I can always improve my level of intelligence. [AC]	6.24 (1.12)		.819
I can get smarter by working hard at learning. [AB]	6.17 (1.15)		.853
If I keep working at something I will get better at it. [BC]	6.37 (.978)		.830
It is possible to change how smart I am. [AC]	5.94 (1.34)		.703
Openness to Challenge and Discovery		.908	
I like to try new things. [AB]	5.62 (1.33)		.833
I am not afraid of trying new things, even if they seem hard. [BC]	5.31 (1.48)		.824
I like to be challenged by new things. [AC]	5.54 (1.39)		.868
I like to try new things, even if I am not very good at them at first. [AB]	5.21(1.48)		.872
Hopeful Purpose		.87	
I am excited about my future. [BC]	5.95 (1.30)		.810
I trust my future will turn out well. [AC]	5.87 (1.31)		.790
My life will make a difference in the world. [AB]	5.53 (1.48)		.797
I am doing things now that will help me achieve my purpose in the world. [BC]	5.88 (1.27)		.813
Prosocial Orientation		.886	
It is important for me to understand how other people feel. [AC]	5.69 (1.27)		.761
I am happy when others succeed. [AB]	6.09 (1.05)		.838
I care about how my decisions affect other people. [BC]	5.92 (1.21)		.848
I can be counted on to help if someone needs me. [AC]	6.28 (.95)		.731
I care about the feelings of my friends. [AB]	6.44 (.89)		.707

*Note: * indicates 2nd Order Factor; ** indicates 3rd Order Factor; λ indicates factor loading; ω indicates Composite Reliability (i.e., Omega); Within 2nd and 3rd order factors, 1st and 2nd order factors act as "items." A, B and C superscript indicate missing data condition.*

Table 11. Descriptive Data from Confirmatory Factor Analyses (Continued)

Factor/Item	M (SD)	ω	λ
Positive Emotionality		.849	
When I want to feel a more positive emotion, I change the way I am thinking about a situation. [BC]	5.11 (1.40)		.857
I control emotions by changing the way I think about the situation I am in. [AC]	5.04 (1.40)		.875
When something upsets me, I try to express how I am feeling rather than pretend I am not upset. [AB]	4.34 (1.70)		.492
When I want to feel less negative emotions, I change the way I am thinking about the situation. [AC]	4.79 (1.44)		.860
Goal Management		.881	
I develop step-by-step plans to reach my goals. [AC]	4.84 (1.58)		.742
If I set goals, I take action to reach them. [AB]	5.45 (1.35)		.895
It is important for me that I reach my goals. [BC]	6.01 (1.19)		.833
I know how to make my plans happen. [AC]	5.38 (1.35)		.756

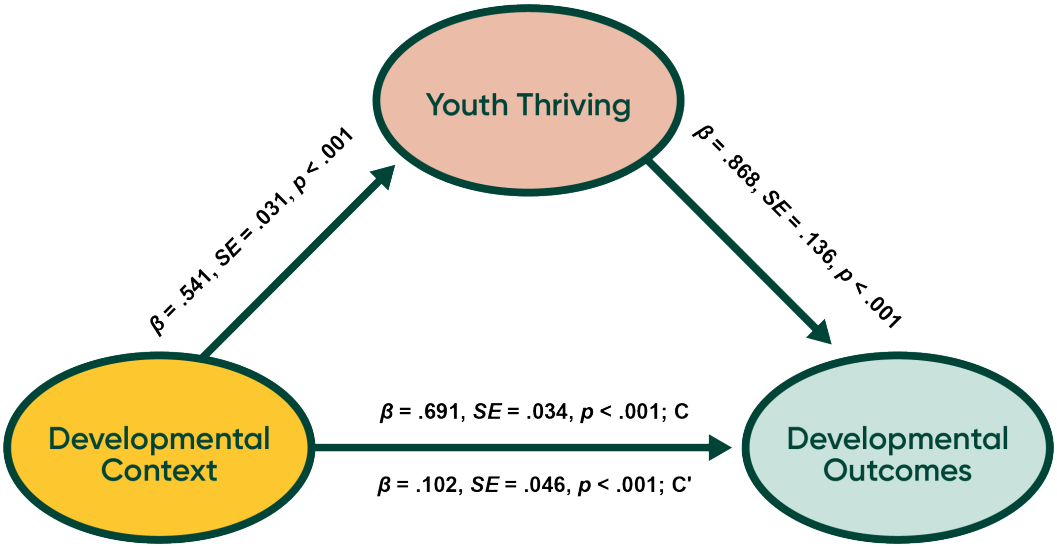
*Note: * indicates 2nd Order Factor; ** indicates 3rd Order Factor; λ indicates factor loading; ω indicates Composite Reliability (i.e., Omega); Within 2nd and 3rd order factors, 1st and 2nd order factors act as "items." A, B, and C superscript indicate missing data condition.*

Testing the 4-H Thriving Model Structure

The structural equation modeling (SEM) results showed a satisfactory model fit, mirroring the confirmatory factor analysis (CFA) findings: $\chi^2(2533) = 10490.467$, $p < .001$, TLI = .921, CFI = .918, RMSEA = .028 (90% CI from .028 to .029), SRMR = .052. The identical model fits between SEM and CFA are attributed to both models having the same number of parameters (further details are available in Appendix B, which includes the Rscript explanation).

In exploring the interrelations among latent variables within a mediation model using lavaan, several significant findings emerged. Firstly, developmental context had a notable direct effect on developmental outcomes without including youth thriving in the model ($B = .691$, $\beta = .569$, $SE = .034$, $p < .001$), termed the "c" path. Additionally, a significant direct effect was observed from developmental context to youth thriving ($B = .643$, $\beta = .541$, $SE = .031$, $p < .001$), and from youth thriving to developmental outcomes ($B = 1.949$, $\beta = .868$, $SE = .136$, $p < .001$). Furthermore, when the direct effect of youth thriving on developmental context was standardized and adjusted for youth thriving ($B = .274$, $\beta = .102$, $SE = .046$, $p < .001$), the results indicated a persistent direct relationship between the independent and dependent variables, suggesting partial mediation. This is also evident in the Sobel test results ($z = 11.79$, $SE = .106$, $p < .001$), indicating a partial mediating effect of youth thriving between developmental context and developmental outcomes. **In other words, youth thriving serves as a partial mediator in the relationship between developmental context and developmental outcomes. This conclusion aligns with the findings of Arnold and Gagnon (2019) which "validated" the scale, and Gagnon et al. (2022) which had a large national sample.**

Figure 5. Mediating Effect of Youth Thriving



Note: First and Second Order Factors, Items excluded for parsimony of presentation. See Table 11 for a comprehensive list of factors and items.

REFERENCES

Arnold, M. E. (2018). From context to outcomes: Adolescent thriving in 4-H Youth Development Programs. *Journal of Human Sciences and Extension, 6*(1), 141-160.

Arnold, M. E. & Gagnon, R. J. (2019). Illuminating the process of youth development: The mediating effects of thriving on youth development program outcomes. *Journal of Human Sciences and Extension, 7*(3), 23-50.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research, 36*(4), 462–494.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

Dashdondov, K., & Kim, M. H. (2021). Mahalanobis distance based multivariate outlier detection to improve performance of hypertension prediction. *Neural Processing Letters.* https://doi.org/10.1007/s11063-021-10663-y

Du, H. & Bentler, P. M. (2022): 40-year-old unbiased distribution free estimator reliably improves SEM statistics for nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*. https://doi.org/10.1080/10705511.2022.2063870

Gagnon, R. J., Arnold, M. E., Lingard, J., & Mullendore, J. (2023). *2022 National 4-H Annual Index Study.* Washington DC: National 4-H Council.

Hayes, A. F. & Coutts, J. J. (2020) Use omega rather than Cronbach's alpha for estimating reliability. But… *Communication Methods and Measures, 14*(1), 1-24.

Henseler, J., Ringle, C. M. & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*, 115–135.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). SEM Tools: *Useful tools for structural equation modeling. R package version 0.5-6.*

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straight-lining: Overview of measurement, comparison of indicators, and effects in mail-web mixed-mode surveys. *Social Science Computer Review, 37*(2), 214–233.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). Routledge.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Yan, T. (2008). Nondifferentiation. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methodology* (pp. 520–521). Newbury Park, CA: Sage.

APPENDIX ONE

Data Screening and Processing

Prior to addressing the research questions, an initial examination of the data was conducted to identify issues such as deviations from normality, outliers, missing values, and respondent eligibility. An evident pattern of "straightlining" was observed, where participants consistently provided identical or nearly identical answers across multiple items. This term, as defined by Yan (2008), refers to a lack of variability in responses, suggesting that participants might not have been thoroughly differentiating their answers to different questions. Such uniformity can negatively impact the accuracy of data analysis. To counter this, the dataset was scrutinized for variability using the 73-item questionnaire as a reference, drawing on the variance analysis methods of Kim et al. (2019) and Yan (2008). Specifically, standard deviations for each participant were calculated for the final quarter of the questionnaire, where inattention typically becomes more prevalent. Here, a standard deviation of zero indicated no variation in responses, an example being a respondent selecting the same score, such as "4" on a scale from 1 to 7, for all 40 items.

To address straightlining, three different criteria were employed:

1. A lenient criterion would exclude participants with a standard deviation of zero, resulting in the removal of 650 participants (13.7% of the sample). This approach could still retain a respondent who selected "4" for 39 items and "3" for one.
2. A more balanced approach, mirroring the methodology of a 2022 report and Arnold & Gagnon (2019), set the exclusion threshold at the lowest standard deviation from the 2019 study (*SD* = .308). This method excluded 760 respondents (16.1% of the sample).
3. A stringent approach used the lower bound of the 95% confidence interval standard deviation from the 2019 study [95% CI (.667, .764)], leading to the exclusion of 1496 participants (31.6% of the sample).

For enhanced accuracy in results, the balanced approach was chosen, leading to a refined dataset of 3,978 potential respondents for further analysis.

The data were further assessed for multivariate outliers using a combination of Mahalanobis distance and the chi-square distribution. This method effectively identifies outliers based on their responses to the entire set of questionnaire items, whether categorical or continuous. This outlier screening indicated that 180 respondents were significantly non-normal ($p < .001$). Consequently, two datasets were prepared: one including and one excluding these outliers. Comparative analysis of model fits, parameter estimates, and descriptive statistics between these datasets showed no significant differences in study results. However, in larger datasets with more variables, stricter outlier criteria may be necessary (e.g., $p < .00001$) (Cohen et al., 2003; Dashdondov & Kim, 2021). Therefore, outliers were maintained in the dataset for subsequent analyses.

The dataset underwent a thorough check for missing data to ascertain the extent of non-responses, such as skipped questions or partially completed questionnaires. When reporting demographic information (such as gender identity), any skipped or missing responses were excluded from the analysis. In contrast, for analyses examining relationships, such as the impact of youth thriving, missing responses were handled using the Full Information Maximum Likelihood (FIML) method, a process elaborated on in the subsequent section on youth thriving. Additionally, the data underwent a screening for state-specific descriptive reporting. In this context, a minimum number of responses were required for a state to qualify for individual state-level reporting. States falling short of this threshold, detailed in Table 1, were still considered in the broader national-level analysis.

APPENDIX TWO

Planned Missing Data Design

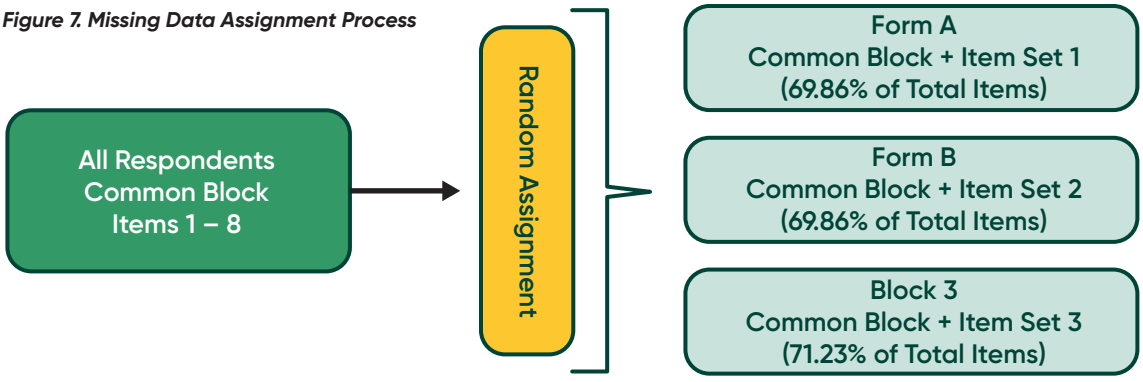
The thriving model was designed to reflect the deep complexity of the processes that inform developmental contexts, developmental outcomes and youth thriving. Correspondingly, the measurement instrument is relatively lengthy with 73 items. To reduce the potential fatigue respondents may encounter with such a lengthy scale, a planned missing data design was employed. In brief, a planned missing data design randomly assigns respondents to a condition where they complete an abbreviated version of a questionnaire. As illustrated in Table 9 and Figure 7, the respondents receive a set of common items, and then are randomly assigned to one of three conditions. This randomization process facilitates the missing data to be missing completely at random (MCAR), which indicates the data are not missing due to a measured participant characteristic. Further, the MCAR conditioning of the data allows for contemporary missing data management techniques. Specifically, in a planned missing data design, the missing data points are then estimated and/or imputed utilizing a technique such as multiple imputation (MI), expectation maximization (EM), or in the case of the present study, Full Information Maximum Likelihood (FIML). With a FIML technique, all available points of data are employed to estimate a respondent's missing values (thus the "Full" in Full Information Maximum Likelihood). The items associated with the three specific missing data conditions are presented in Table 9.

Table 9. Planned Missingness Design

	Common Items (8-Items) + (Item Set 1, 2 or 3)	Item Set 1 (43-Items)*	Item Set 2 (43-Items)*	Item Set 3 (44-Items)*
Form A	X	X	X	•
Form B	X	•	X	X
Form C	X	X	•	X

Note: All respondents received common items; X indicates respondent received item; • indicates planned missing data; * is total number of items in set excluding common items.

Figure 7. Missing Data Assignment Process



Data Diagnostics and Analytic Plan

As the framework used to test the measurement properties of the scale and study hypotheses assumes multivariate normality, the data were screened for multivariate normality utilizing the MissMech package (version 1.0.2; Jamshidian et al., 2014). This analysis indicated the data were multivariate non-normal (Hawkins's test $p < .001$; Anderson Darling k-sample test, $p = .035$). Because of this, robust estimation techniques were employed to adjust for the violation of normality necessary for maximum likelihood (ML) analyses. Specifically, maximum likelihood estimation with robust (i.e., MLR) standard errors were utilized as they reduce potential Type 1 error and/or misinterpretation of model fit and parameter estimates (Du & Bentler, 2022). The data were then screened for multivariate outliers utilizing a combination of Mahalanobis distance and the chi square distribution. This analysis suggested 102 respondents were significantly ($p < .001$) non-normal. As such, two models were produced for both the measurement and hypotheses testing, one with and one without the outliers in the data set. Comparison of model fits with the two data sets did not indicate statistically meaningful differences across the model fit indices. As such the outliers were retained.

The measurement properties and hypotheses were tested through confirmatory factor analysis (CFA) and structural equation modelling (SEM) utilizing the lavaan package (version 0.6-12; Rosseel, 2012) and the semTools package (version 0.5-6, Jorgensen et al., 2022) in R (version 4.2.2). The acceptability of model fit for both the CFA and SEM were examined utilizing a robust version of the Root Mean Squared Error of Approximation (RMSEA), where values closer to zero indicate a model fit that reflects the properties of the data (e.g., RMSEA < .070). Similarly, the standardized root mean square residual (SRMR) was utilized, which assesses the differences between observed correlations and hypothesized/predicted correlations, and values closer to zero also indicate better model fit (e.g., SRMR < .100) (Kline, 2016). Additionally, the Tucker Lewis Index (TLI) and Comparative Fit Index (CFI) were employed, as they assess the degree to which the specified model is an improvement over a null/non-specified model. In both the TLI and CFI, scores closer to one are preferable (e.g., TLI > .900) (Loehlin & Beaujean, 2017). Model fit criteria were not assessed on arbitrary fixed cutoff scores (e.g., RMSEA = .071 is unacceptable versus RMSEA = .069 is acceptable), rather they were assessed based upon the model complexity and past model performance (Chen et al., 2008).

In addition to model fit, the measurement model was also examined for convergent and discriminant validity. Within the tests of convergent validity, the factor loadings were assessed (e.g., the degree to which the theorized factor reflects the specified items), where scores closer to one suggest a stronger predictive influence of the factor on the item, and lower scores (i.e., $\lambda < .400$) suggest the item may need to be respecified within the measurement model. The convergent validity was also assessed by employing the McDonald's omega (ω), a measure of internal consistency that performs better than Cronbach's Alpha (i.e., α), as it does not have the generally unmet threshold of tau equivalence informing α (Hayes & Coutts, 2020). Additionally, the Average Variance Extracted (AVE) levels were examined to ensure the factors were accounting for more variance than error (i.e., AVE > .500). The discriminant validity of the scales was also examined to ensure the scales were reflecting distinct constructs. The between factor correlations were examined, where lower values indicate the factors are accounting for unique variance (i.e., $r < .700$). Specifically, the square root of the AVE was examined to ensure the variables were accounting for more unique variance than sharing with other factors in the model ($\sqrt{AVE} \geq r$). Similarly, the heterotrait-monotrait ratio of correlations (HTMT) was examined, where values less than one indicate discriminant validity (i.e., HTMT ratio < 1.00) (Henseler et al., 2015).