


# I Am Whatever I Say I Am: The Salient Identity Content of U.S. Adolescents

Sara K. Johnson , Katharine Odjakjian, and Yerin Park  
*Tufts University*

Research about identity development has focused primarily on researcher-chosen domains or overlooked content entirely. To investigate the content that is salient to adolescents, we analyzed responses to a shortened Twenty Statements Test (ten answers to the question, “Who am I?”) from 415 adolescents in the northeastern United States ( $M_{\text{age}} = 13.59$  years; 63.7% girls, about 50% identified as White/European American). Inductive content analysis identified four Content codes (what the participants wrote: Personal, Social categories, Relationships, Self-evaluation) and two Structure codes (how they wrote their statements: Qualifiers and Verb Tense). Content codes appeared in eight patterns, and there were between-group differences in content and patterns. Results expand our understanding of adolescents’ identity content and demonstrate its complexity.

Key words: adolescence – identity content – Twenty Statements Test

Cultivating a sense of identity is one of the most important parts of growing up, and many adolescents ponder the fundamental identity questions of “Who Am I?” and “Who do I want to become?” In contemplating these questions, they may consider many topics, issues, and aspects of themselves ranging from social roles, group memberships, and relationships, to personality characteristics, interests, and potential occupations. These aspects can be described as identity content: the *elements* that make up identity (Galliher, McLean, & Syed, 2017). Knowing about identity content (e.g., a future occupation) is equally important as knowing about identity processes (e.g., exploration about potential future occupations). Content, however, has not been a major focus of developmental research on identity. Most studies have addressed content in an implicit and top-down way; typically, questions about processes (e.g., exploration) have been asked within a few researcher-chosen content areas (e.g., education) that may not be salient to all participants equally (Waterman, 2015). There has been little research wherein young people are able to describe the content most salient to them (*cf.*, Karaś, 2015; McLean, Syed, Yoder, & Greenhoot, 2016), even though this type of information is critical for a complete understanding of identity and its development (Galliher et al., 2017). To address

this gap, we explored how identity content was described among a sample of adolescents from the northeastern United States.

## Conceptualizing Identity Content

Content has been addressed in many fields (e.g., developmental psychology, social psychology), but in diverse ways, and even within fields, it has been conceptualized differently. The most common definitions are quite broad: Content is “what the identity looks like (McLean & Syed, 2015, p. 4), “what the identities are” and “the description of who one is” (Syed & McLean, 2015, pp. 568–570), and “what is developing” (McLean, Syed, & Shucard, 2016, p. 356). Galliher et al. (2017) proposed that a singular definition of content may not even be desirable; instead, they offered a multilevel framework with levels of content including culture, social roles, domains, and everyday experiences. These conceptualizations of identity content informed our perspective for the analyses presented here.

Our definition of content is based on our conceptualization of identity itself, specifically its use in developmental research. We view identity as a person’s sense of sameness and continuity as organized by the self and recognized by others. Identity content, then, consists of the elements on which that sense is based (i.e., self-characteristics that give one a sense of sameness and continuity). We acknowledge that this definition is broad and overlaps with related concepts. Literature on aspects of the self-system and its development is varied and offers overlapping (and sometimes contradictory)

---

This research was sponsored in part by a grant from the Templeton Religion Trust to Jacqueline V. Lerner (Principal Investigator) and Sara K. Johnson (Co-Principal Investigator).

Requests for reprints should be sent to Sara K. Johnson, Evans Family Assistant Professor, Eliot-Pearson Department of Child Study and Human Development, Tufts University, 105 College Avenue, Medford, MA 02155. E-mail: S.johnson@tufts.edu

definitions and operationalizations of terms, including self-concept and identity. Accordingly, what we define as content is also addressed within research on other aspects of the self (e.g., self-concept; Harter, 2015). In the analyses presented here, our intention was to explore what information arises when adolescents are directly asked the basic identity question of “Who Am I?” We assumed that if an adolescent wrote something as an answer to “Who Am I?” then the response *most likely* had some relevance to identity content (as others have also assumed; e.g., Hihara, Umemura, & Sugimura, 2019; Naudé, 2020).

Of course, we could not be certain that all responses to this question were considered by adolescents themselves to be part of their identity content, and many responses may reflect multiple self-relevant constructs (especially given the overlap between self-related constructs). Adolescents may view some parts of themselves as important to their own understanding of who they are as a person, whereas other aspects may simply be self-description or reflective of how others view them. It may not be possible to precisely determine whether a characteristic that individuals use to describe themselves could be classified as identity content without follow-up questions of the type conducted during an in-depth interview. For example, a characteristic that appears to be simply self-description (e.g., I am tall) may be a crucial aspect of identity for someone because of the meanings attached to the characteristic.

Given that not much is known about content, we chose to err on the side of inclusivity rather than preemptively (and perhaps prematurely) eliminating specific responses based on preconceived ideas about what content should be. As part of our inclusive approach, we did not classify adolescents’ responses in a binary manner as *definitively* reflective of identity content as opposed to any other self-relevant construct, and there were only a small number of statements that we coded as not being content-related based on our definition (see the Method section). We believed this approach was important given the lack of research on the identity content that adolescents themselves report.

### Empirical Research Regarding Identity Content

*Researcher-driven and implicit approaches to content.* Most identity development research has addressed content in a top-down (i.e., researcher-driven) and implicit way through the concept of domains—broad categories—of identity. Beginning

with the identity status interview (Marcia, 1966), eight domains have been used (e.g., Balistreri, Busch-Rossnagel, & Geisinger, 1995): Religion, occupation, politics, and values are often grouped together as ideological, whereas romance, friendship, family, and gender roles have typically been grouped as interpersonal.

These domains continue to be referenced in contemporary questionnaire measures of identity processes (e.g., exploration and commitment). For example, the Utrecht Management of Identity Commitments Scale (Crocetti, Rubini, & Meeus, 2008) includes questions about processes with reference to relationships and education. The domains themselves, however, typically do not play a role in analyses because scores on the items are used to make general conclusions about processes. For example, in-depth exploration in education may be used to represent general exploration in depth, with the assumption that identity processes are generalizable across domains (although some researchers have questioned this assumption; Goossens, 2001; Karaś, 2015; McLean, Syed, & Shucard, 2016). A similar approach has been taken in research on identity related to group memberships (i.e., social identity), such as gender identity (e.g., Bussey, 2011) and racial-ethnic identity (Umaña-Taylor, 2011). Participants are often asked how much they think and learn about (i.e., explore) these group memberships, and how committed they feel to them, rather than the specific content or meaning of the group memberships.

These methods provide considerable information about processes (as they were designed to do), but they provide limited information about content. Some information can potentially be inferred, given that low levels of exploration or commitment may indicate a low level of salience of the domain in which the questions are asked. For example, low exploration in the educational domain could indicate that education is less important to the identity of the respondent. That may not be true for all participants, however (Waterman, 2015), and it is often unclear what the content itself looks like.

Research about narrative identity processes (e.g., autobiographical reasoning, meaning-making) has also incorporated the domain approach, but the focus has also been on processes rather than domains. Fewer studies have addressed the role of content in participants’ narrative processing. For example, McLean, Syed, and Shucard (2016) asked participants to complete narrative prompts within the usual eight domains. They noted that those eight domains were generally salient overall but

with “notable” variability” (p. 361), and some domains “spilled” into others. For example, domains typically classified as interpersonal (e.g., family) appeared frequently in narratives in other domains, including most of the ideological ones. The authors interpret those findings as indication that some contents (e.g., family) serve an integrating purpose across domains. The approach of McLean, Syed, Yoder, et al. (2016) has the advantages of providing some indication of which domains may be more relevant to participants, and the ways in which specific content domains may overlap. However, participants were not asked directly about their identity content, but rather to provide narratives in content areas preselected by the researchers.

Some research on social identity development has included questions about the centrality of domains, such as racial or ethnic identification (Brittian, Umaña-Taylor, & Derlan, 2013). This research has demonstrated that some domains are more central for some people than others (e.g., Meca et al., 2015), and these differences are related to societal contexts. For example, people from groups that are marginalized within a specific context (e.g., in the United States, people of color) typically rate that aspect of their identity as more central compared with people who are members of the majority or power-holding group (Bombay, Matheson, & Anisman, 2010). This research points to interindividual variation in the salience of identity content elements, but the participants are again responding only to one domain preselected by the researchers.

Similarly, there is considerable research on “niche” content. This literature includes (among many) volunteer identity (Thoits, 2013), vegetarian identity (Rosenfeld & Burrow, 2017), athletic identity (Martin & Horn, 2013), and teacher identity (McNaughton & Billot, 2016). These studies usually originate from an interest of the researcher, however, so it is unclear how salient they are generally.

A few researchers have investigated the content people spontaneously reference when they are answering other types of questions. For example, McLean, Syed, Yoder, et al. (2016) coded the content that young adults included in personal narratives. The eight traditional domains indeed appeared, but not all domains were equally salient to all participants. Furthermore, inductive coding showed contents beyond these eight (e.g., existential). This approach uses an implicit way to get information about content (e.g., in response to

questions not directly about content) and shows the broader domains that are salient to participants insofar as they are reflected in the provided narratives. However, participants only supply a few elements at a time, which limits information about the larger number of elements of identity content that are meaningful.

*Participant-driven and explicit approaches to content.* A few other researchers have asked participants more explicitly about their identity content. In qualitative interviews, Karaś (2015) asked young adults to answer the question of “Who am I?” The contents mentioned included (from most to least often) personality characteristics, past experiences, family, friends and acquaintances, worldview, hobbies and interests, aims and plans for the future, and occupation. However, in-depth interviews are likely too labor-intensive for large-scale studies.

One option for collecting participant-supplied information about identity content in surveys is the Twenty Statements Test (TST; Kuhn & McPartland, 1954). The TST is a sentence-completion task in which participants complete up to 20 sentences that begin with “I am...” Over its more than 50-year history, the TST has been used to examine many aspects of the self with different samples and administration methods, including with adolescents (Ferguson, Ferguson, & Ferguson, 2017) and adults (Bayerl, Horton, & Jacobs, 2018). Researchers have used the TST to compare self-descriptions of people from different professions (e.g., Bayerl et al., 2018), and analyzed responses for specific content (e.g., Lumma, Böckler, Vrticka, & Singer, 2017).

A few studies using the TST have been explicitly labeled by the authors as investigations of identity content. Hihara et al. (2019) coded the valence of Japanese young adults' TST responses as positive (socioculturally desirable identity content in the Japanese sociocultural context), negative (socioculturally undesirable identity content), or neutral (neither desirable nor undesirable). A follow-up study (Hihara, Sugimura, Umemura, Iwasa, & Syed, 2021) showed that these content valences were related to adaptive and maladaptive behaviors. Naudé (2020) coded TST responses of South African adolescents and young adults, both deductively (for desirability/undesirability and independence/interdependence) and inductively (for content types). The domain-specific content identified included social qualities; hobbies, sport, and interests; and values. Adolescents provided more

statements with intellectual/ cognitive and physical qualities, whereas young adults included more social qualities such as altruism.

### THE CURRENT ANALYSES

Research about identity processes has either overlooked aspects of content or focused within a few researcher-chosen domains. Accordingly, information about identity content—particularly the types of content that participants identify as salient—is needed for a more complete and ecologically valid picture of identity (Galliher et al., 2017). Here, we used the TST to explore the identity content of adolescents from the north-eastern United States. We had three overarching research questions.

The first question was: What kinds of identity content are salient to adolescents in the United States, and which contents are reported most often? We expected that the eight domains used in prior identity research would be included in the responses, but that the responses would reflect more than just those eight categories. The second question was: Do these types of content occur in patterns? We considered these analyses entirely exploratory and did not have *a priori* expectations. Finally, the third question was: Do the overall types and patterns of content differ between groups of adolescents? We explored differences related to age, gender, and racial-ethnic identification. For age, we anticipated that older adolescents might have a larger variety of types of codes applied to their statements, given that identity complexity increases with age. For gender, we expected that girls might be more likely to include gender identification, given the lower status of girls in the US context. Finally, for racial-ethnic identification, we expected that adolescents of color would be more likely to include their racial and/or ethnic identification in their statements than White youth, given the lower status of youth of color in the United States. We did not have specific expectations about between-group differences in identity content patterns (given that we did not have expectations about the patterns themselves).

## METHOD

### Research Context

We used data from the second wave of the Connecting Adolescents' Beliefs and Behaviors (CABB) Study (Johnson et al., 2016; Lerner, Wong, Weiner,

& Johnson, 2021). CABB was a three-year (2015–2017) mixed-method, multireporter study of the positive development of adolescents in fifth through twelfth grades. CABB includes data from youth, their parents, and adults at their schools (e.g., teachers). We used only the data collected from youth.

### Procedures

The CABB research team recruited adolescents through a two-step procedure by first recruiting schools and then recruiting young people from within those schools. They approached 37 schools in Massachusetts and Connecticut; fifteen agreed to participate. Due to scheduling challenges (e.g., severe weather during the winter of 2015 in New England), data were collected at only twelve of those schools (eight Catholic/parochial and four public). Schools received a \$200 gift card. Once school administrators agreed to participate in the study, the research team provided them with packets of consent forms to send home with students. The students' parents/guardians provided consent for their children either online or by returning a paper copy of the consent form. At a time that was convenient for the school, trained research staff collected data from students. Students provided written assent and then completed a paper-and-pencil survey, which took about an hour. Students received a \$15 gift card.

### Participants

Of the 419 adolescents from the second wave of the CABB Study, four did not write any TST responses. The remaining 415 comprised the analytic sample. They were in 6th through 12th grades ( $M_{\text{age}} = 13.58$  years,  $SD = 1.90$ ; 13 participants had no age information). Most identified as girls ( $n = 266$ ; 65.0%), 143 (35.0%) identified as boys, one (0.20%) participant did not provide gender information, and five (1.2%) chose "Other" for gender identity. Two hundred and five (49.4%) identified as White/European American, 62 (14.9%) as Black/African American/of African descent, 48 (11.6%) as Hispanic/Latino/Latina, 46 (11.1%) as multiracial, 37 (8.9%) as Asian or Asian American, 8 (1.9%) as Other, 5 (1.2%) as Arab American or Middle Eastern, and 1 (0.2%) as Pacific Islander; 3 (0.7%) did not respond. Information about parental education, obtained through the parent survey, was available for 201 participants (48.4%). Among them, 15 parents (7.5%) had completed high school or

equivalent, 31 (15.4%) had attended or graduated from a vocational/trade school or community college, 84 (41.8%) had attended or graduated from a four-year institution, and 71 had attended graduate school or received a graduate degree (35.3%).

The demographic variables were related in several ways. Girls ( $M = 13.77$  years) were about half a year older than boys on average ( $M = 13.25$ ),  $t(394) = -2.58$ ,  $p = .01$ . Participants who identified as White/European American ( $M = 13.18$ ) or multiracial ( $M = 13.15$ ) were younger on average than participants who identified as Black/African American/of African Descent ( $M = 14.18$ ), Hispanic/Latino/Latina ( $M = 14.07$ ), or Asian/Asian American ( $M = 14.05$ ),  $F(4, 384) = 5.815$ ,  $p < .001$ . Gender proportions were similar across racial groups,  $\chi^2(4, n = 392) = 2.72$ ,  $p = .61$ .

## Measures

The CABB Study used a shortened version of the TST (Kuhn & McPartland, 1954; see also Rohall, 2018). The prompt was: *Please give ten different statements to answer the question, "Who am I?" Give these answers as if you were saying them to yourself, not to somebody else.* Participants were given ten prompts starting with "I am" followed by a blank space.

## Analysis Plan

**Research Question 1: Types and prevalence of content.** To describe and categorize adolescents' responses to the TST, we conducted an inductive content analysis (Hsieh & Shannon, 2005) at the statement level. The second and third authors coded the statements; the first author served as the auditor by reviewing the statements and associated codes and discussing them with the coding team. Accordingly, all statements were double-coded (by the second and third authors) and reviewed by the first author, using the following procedure. Each statement could receive more than one code.

The initial step was conducted by the first and second authors. They created a preliminary codebook with three categories (personal, social, and relational) based on identity development literature and reviewing the statements of 10 participants. The second author then coded the statements of 50 more participants and noted new categories, categories that could be linked together as main codes, and codes that needed to be altered to better match the data. She discussed these notes with the first author, and we made changes to the codebook, as well as to how some statements were coded, based

on these discussions. A major initial change was the addition of Structure codes (see Results). When new codes were added, the second author re-coded all statements to reflect the updated codebook. The second author coded the remainder of the statements in sets of 50 participants. After each set, the first author reviewed the codes, and they discussed areas of confusion and potential codebook revisions.

In the second phase, the third author joined the coding team. After familiarizing herself with the codebook, she independently coded the statements of 10 participants (without consulting the second author's codes). She then reviewed her coding with the second author, and they addressed differences through discussion and consultation with the first author; minor clarifications to the codebook were made, and a few codes were added. The process was repeated with sets of 20 and then 50 participants with very few changes to the codebook. Once all statements had been double-coded, the third author reviewed all statements to ensure that they were coded consistently with the final codebook.

In the third step, we evaluated the codes to determine whether they all indicated what we considered to be identity content. Given that we wanted to be inclusive, we chose to categorize all codes as related to content except for the code we named Temporary States. This code was applied to 43 statements that included only references to something the participant was doing at that moment (e.g., "I am holding a pen right now" or "I am filling out a survey") related to the situational context (i.e., research participation) and thus unlikely to provide a sense of sameness and continuity to participants.

To determine content prevalence, we computed descriptive statistics at the statement and participant levels. At the statement level (i.e., pooling statements across participants), we computed frequencies of the main and subcodes. At the participant level, we computed descriptive statistics about the number of total codes, main codes, and subcodes given to a participants' set of statements.

**Research Question 2: Patterns of content.** The analysis strategy for this question was not determined in advance because it depended on the analysis for research question 1. Accordingly, we describe it in more detail in the Results section.

**Research Question 3: Between-group differences in content.** Our final aim was to explore between-group differences in content (at both the statement

and pattern levels) based on three demographic variables. For analyses involving racial-ethnic identification, we included students from the largest five groups (White, Black/African American/of African descent, Asian or Asian American, Hispanic/Latino/Latina, and multiracial). To test for differences in the presence of codes between racial-ethnic and gender groups, we used chi-square analyses. To test whether the number of codes differed between racial and gender groups, we used analyses of variance. For similar analyses related to age, we used regression analyses so that we could also test for nonlinear relationships; in these regressions, we predicted the number of codes based on both linear and quadratic age terms. Because girls and racial-ethnic minority participants were older on average, we also conducted exploratory follow-up analyses for age separately by gender and in the five largest racial/ethnic groups. Finally, we used chi-square analyses and analyses of variance to test whether participants' code patterns (identified in research question 2) were related to racial-ethnic identification, gender identification, or age.

*Post hoc analyses of word count.* Because these analyses involved frequencies of information provided in open-ended responses, we also conducted post hoc supplemental analyses of word count (at both the statement and participant levels) and its relation to the type, number, and pattern of codes. In the Results section, we include descriptive information about word count and highlight the few cases where it played a notable role. The remainder of word count analyses are in Appendix S1.

## RESULTS

The 415 participants provided a total of 4,043 statements. Of these, 15 were coded as uninterpretable (e.g., nonsensical letters) and 43 as not related to identity content (i.e., related to Temporary States). The remaining 3,985 statements were coded using procedures described above.

Most statements ( $k = 2,367$ ; 59.4%) were only one word; 479 (12.0%) included two words, 524 (13.1%) had three, 200 (5.0%) had four, and 137 (3.4%) had five; the remainder ( $k = 278$ ; 6.9%) included six or more words, with a maximum of 20 ( $k = 3$ ; 0.10%). The average was 2.16 words ( $SD = 2.04$ ).

Most adolescents ( $n = 343$ ; 82.7%) wrote 10 statements, 38 provided nine, 14 wrote eight, and 20

provided seven or fewer. The average was 9.6 statements ( $SD = 1.19$ ). Average word count (across all statements of a participant) ranged from 1 ( $n = 69$ ; 16.6%) to 15.2 ( $n = 1$ ; 0.20%),  $M = 2.16$  ( $SD = 1.58$ ).

### Research Question 1: Types and Prevalence of Content

Statements were categorized into two types: Content referred to information (what participants wrote about), and Structure referred to features of the writing (how they wrote about it). Both Content and Structure had main codes; most of the main codes had subcodes. Table 1 shows Content codes, and Table 2 shows Structure codes; both tables include definitions, subcodes, and frequencies.

Most statements ( $k = 3,029$ ; 76.0%) had only one code (of any type); 783 statements (19.6%) had two, 162 (4.1%) had three, nine (0.2%) had four, one had five, and one had six codes. At the participant level, the number of codes for a set of statements ranged from 1 ( $n = 1$ ; 0.2%) to 32 ( $n = 1$ ; 0.2%), with the largest proportion of participants ( $n = 107$ ; 25.8%) having 10 codes applied to their set.

Overall, statements with more codes had more words on average,  $F(2, 3982) = 1457.8$ ,  $p < .001$ . Average word counts were 1.47 for statements with one code (range = 1–12), 3.84 words for two codes (range = 1–14), and 6.59 for three codes (range = 2–20). Yet, word count and number of codes were not redundant: Of the 2,367 single-word statements, 42 (1.7%) had two codes, and 704 (43.5%) of the 1,618 statements with more than one word received only one code.

*Content codes.* All 3,985 statements were coded with at least one Content code. Most statements ( $k = 3,483$ ; 87.4%) had only one; 479 statements (12.0%) had two Content codes (of any subtype), 19 (0.5%) had three, and 4 statements (0.1%) had four. All participants had at least one Content code applied to their set of statements. The total number ranged from 1 ( $n = 2$ ; 0.5%) to 21 ( $n = 1$ ; 0.3%). The largest percentage ( $n = 169$ ; 40.7%) had 10 Content codes applied to their statement set.

We identified four main Content codes: 1. Personal, 2. Social Categories, 3. Relationships, and 4. Self-evaluation. See Table 1 for definitions, examples, and frequencies of the subcodes.

*Personal.* Personal codes were the most prevalent ( $k = 3,611$ ; 91% of Content codes). We coded

TABLE 1  
Main and Subcodes for Content, Including Code Descriptions and Examples, and Number of Statements with Each Code

Code	k	Description	Example
<b>Personal</b>			
Personal characteristics	3,066	A general description of themselves	I am funny
Hobby	263	An activity that they take part in	I am musician
Physical features	190	physical characteristics or appearance	I am beautiful
Emotions	178	Participant mentions feeling an emotion	I am happy
Student	96	Status as a student/anything related to schooling	I am a high school student
Likes	96	Something they like/love, or are interested in	I like to play with everyone
Self-reference	90	The identifier is solely a reference to the self	I am me
Age	39	Their age; or more general age references	I am 14 years old; I am young
Desires	15	Something they want	I am wanting to make a change
Fan	14	Of a team, celebrity, etc.	I am a Golden State Warriors fan
Religious affiliation	12	Description of religious affiliation	I am Catholic
Values	11	Something the participant values or believes in	I am a person who loves diversity
Dislikes	9	Something they say they don't like	I am person who doesn't like drama
Illness	9	Having an illness, physical or mental	I am bipolar
School	9	The name of their school	I am a student at [name of school]
Grade	7	Their grade level	I am in 7th grade
Personhood	7	Personhood, or simply being a person	I am a person; I am human
Occupation	6	Career aspirations	I am hope to be a teacher
Political affiliation	3	Description of political affiliation	I am conservative
Group membership	3	Mention of a group or team they are part of	I am in an anime club
<b>Social categories</b>			
Gender identity	70	Any mention of their gender	I am a respectful young man
National Identity	37	Country they identify with	I am love my country U.S.A.
Ethnic identity	35	Any mention of their ethnicity	I am Haitian
Racial identity	20	Any mention of their race	I am Black
Sexual orientation	7	Any mention of their sexual orientation	I am a lesbian
<b>Relationships</b>			
Friendship	110	Mention of friendship	I am a very good friend
Familial	96	Mention of any family member	I respect my parents
Spiritual	17	Relationship with a spiritual/religious figure	I am grateful for God
General	15	General description of relationships	I am a member of different communities
Animal	7	Mention of pet or relationships with animals	I am grateful for my dog
Romantic	3	Mention of a romantic relationship	I am in a relationship
Classmate	3	Mention of classmates	I am a classmate
<b>Self-evaluation</b>			
Positive	388	Evaluation of self or a characteristic positively	I am the best; I am good at math
Negative	20	Evaluation of self or a characteristic positively	I am not always happy with myself
Neutral	15	Evaluation of self is not clearly positive or negative	I am all in all an ok person

statements as Personal when they included information about the self, independent of relationships or social categories. The most common subcode was general personal characteristics ( $k = 2,745$ ; 76% of statements coded as Personal), such as “funny” and “educated on social issues.”

*Self-evaluation.* Self-evaluation codes were applied to 400 statements (9% of statements with Content codes). We coded statements into this category when they included an evaluation of the whole self or part of the self (e.g., skills or qualities). Most statements had a positive valence ( $k = 364$ ), but others were negative ( $k = 20$ ) or neutral ( $k = 16$ ).

*Relationships.* Elements of Relationships were coded in 222 statements (6% of statements with Content codes). We coded statements in this category if another person or living being was mentioned (e.g., friends, family, romantic partners, pets).

*Social categories.* The Social Category code was applied to 133 statements (3% of statements with Content codes). Statements were coded this way when they included a social group (e.g., gender, nationality, ethnicity, race, sexual orientation).

*Content codes and word count.* Overall, statements with more words had more Content codes. This pattern was consistent across all Content

TABLE 2  
Main and Subcodes for Structure, Including Code Descriptions and Examples, and Number of Statements with Each Code

<i>Code</i>	<i>k</i>	<i>Description</i>	<i>Example</i>
Qualifier			
Degree	220	How much a characteristic applies	I am funny (Kinda)
Conditional	146	Conditions under which something applies	I am cold-hearted (toward boys)
Negation	82	Explicitly *NOT* a specific characteristic	I am not selfish
Positive	38	Reflect on a specific characteristic positively	I am weird (in a good way)
Willingness	27	They say they are willing to do something	I am willing to take my time to help
Comparative	18	How they compare to others on a characteristic	I am a kid that is more mature than others
Self-comparison	13	Comparing to a potential state of self or past self	I am the best I can be
Presentation	7	Characteristic that conflicts with self-view	I am pretending to be happy
Negative	6	Reflect on a specific characteristic negatively	I am over protective
Uncertainty	3	Uncertainty about a specific characteristic	I am unsure if I am depressed
Tense			
State of becoming	63	Current actions in the active voice, working toward something like an ideal self	I am becoming comfortable in my own skin
Future-oriented	32	References the future	I am going to chase after my dreams no matter what
Past tense	3	References the past (but not the present)	I am shy didn't like to talk in class
Continuation	4	The characteristic applied to them before and still does	I am still insecure about my body

subcodes (i.e., word count was only related to the number of codes and not the type). Please see Appendix S1 for full details.

**Structure codes.** We coded statements with respect to their structure if there were features that were not about content specifically (e.g., could apply to any type of content) but instead were modifiers of the content or the way the participant wrote the statement. Most statements ( $k = 3,442$ ; 86.4%) had no Structure codes; 475 (11.9%) were coded with one, 65 (1.6%) were coded with two, and three statements (0.1%) had three Structure codes. Slightly less than half of the participants ( $n = 193$ ; 46.5%) did not have any Structure codes for their set of statements. Among participants with Structure codes, the number ranged from 1 ( $n = 84$ ; 20.2%) to 13 ( $n = 1$ ; 0.2%),  $M = 1.48$  ( $SD = 2.14$ ). The two main codes were Qualifier and Tense. See Table 2 for definitions, examples, and frequencies of the subcodes.

**Qualifier.** Four hundred and seventy-four statements (87.0% of statements with Structure codes and 12.0% of all statements) contained a qualifier in which participants gave additional information about the content they had written about, such as how much or under what situations the descriptions applied to them. For example, one participant wrote "I am somewhat athletic."

**Tense.** Most statements were written in the present tense of the verb "to be"; others had different types of verbs or indications of time perspective; we coded these statements as having a different verb tense. An example statement is, "I am becoming comfortable in my own skin." This code was applied to 90 statements (19% of statements with Structure codes and 2% of all coded statements).

**Structure codes and word count.** Overall, statements with more words had more Structure codes. This pattern was consistent across both Structure subcodes (i.e., word count was only related to the number of codes and not the type). Please see Appendix S1 for full details.

**Relation between structure and content codes.** Structure codes were more prevalent in statements with multiple Content codes,  $\chi^2$  (1,  $k = 3,985$ ) = 8.15,  $p = .004$ . Of statements with one Content code, 13% ( $k = 454$  of 3,483) had one or more Structure codes, whereas 17% ( $k = 89$  of 502) of statements with two or more Content codes had at least one Structure code.

Beyond this overall relationship, Structure codes were more or less likely to be present for certain types of Content codes. For Personal codes, there was a positive association,  $\chi^2$  (2,  $k = 3,985$ ) = 246.09,  $p < .001$ : Among statements with zero Personal codes or one Personal code, only about one in ten (12.0% and 13.0%, respectively) had a Structure code. However, among statements with two or more Personal codes, over one-third (33.9%) had a Structure code. A similar pattern was observed for Relationship codes,  $\chi^2$  (1,  $k = 3,985$ ) = 15.18,  $p < .001$ . Among statements with no Relationship codes, 13.0% had a Structure code, whereas 21.0% of statements with at least one Relationships code *also* had at least one Structure code. The opposite pattern was found for Self-Evaluation codes,  $\chi^2$  (1,  $k = 3,985$ ) = 9.93,  $p = .002$ . Among statements with no Self-Evaluation codes, 14% had at least one Structure code, whereas only 8% of statements with a Self-Evaluation code also had a Structure code. For Social Categories codes, there was no relationship between whether a statement had a Social Category code and whether it had a Structure code,  $\chi^2$  (1,  $k = 3,985$ ) = 0.30,  $p = .59$ .

**Structure codes, content codes, and word count.** Finally, we examined the role of word count in the relationship between Content and Structure codes. Overall, average word count was higher in statements with multiple codes, but there were no clear differences based on whether those codes were Content or Structure (i.e., word count was related to number of codes rather than type).

## Research Question 2: Patterns of Content

We next explored how codes occurred together (i.e., patterns). We considered these analyses exploratory (as they were based on the results from research question one), so we did not have an analysis strategy developed *a priori*. Accordingly, we sought a way to capture the number and nature of each type of code within participants' sets of statements that would balance between parsimony and complexity.

The strategy we chose was to create binary variables for each participant indicating the *presence* (yes/no) of the four Content codes. Cross-tabulation of these variables resulted in 16 possible patterns (e.g., only Personal codes; at least one Personal code and at least one Relationships code). Only 8 of these possible patterns were present

(because all participants had at least one Personal code): 128 participants had only Personal codes (30.1%), 99 (23.9%) had Personal and Self-Evaluation, 66 (15.9%) had Personal, Relationships, and Self-Evaluation, 33 (7.9%) had Personal and Relationships, 33 (7.9%) had Personal and Social, 24 (5.8%) had all four types, 19 (4.6%) had Personal, Social, and Self-Evaluation, and 13 (3.1%) had Personal, Social, and Relationships codes. Overall, 30% of participants had statement sets with only one type of Content code; 39.8% had two different Content codes, 23.5% had three different types, and 5.8% had four. See Table 3 for example sets of statements for each of the patterns.

We next evaluated the relationship between the Content code patterns and Structure codes. As described earlier, only a small percentage of statements had at least one Structure code, and Structure codes were not equally distributed across the patterns,  $\chi^2(7, n = 415) = 27.75, p < .001$ . Structure codes were more likely to be present in patterns that included multiple other types of codes. Participants with the most common code pattern (Personal codes only;  $n = 128$ ) were the least likely to have at least one Structure code applied to their statements ( $n = 46$ ; 35.9%). Structure codes were slightly more common (present among about half) for participants with one of the three pattern types consisting of two codes: Personal and Relationships only ( $n = 18$  of 33; 54.5%); Personal and Social only ( $n = 18$  of 33; 54.5%); and Personal and Self-Evaluation only ( $n = 57$  of 99; 57.6%). Structure codes were most common within patterns that had at least three types of codes: Personal, Social Categories, Relationships ( $n = 8$  of 13; 61.5% had one or more Structure codes); Personal, Relationships, and Self-Evaluation ( $n = 46$  of 66; 69.7%); Personal, Social Categories, Self-Evaluation ( $n = 14$  of 19; 73.7%); and Personal, Social Categories, Relationships, Self-Evaluation ( $n = 15$  of 24; 62.5%).

Finally, we investigated word count differences between the patterns. Participants with only Personal codes had the lowest average word count, but it was not statistically significantly different from the word count for two other patterns (which both involved more than one code type). Moreover, participants with all four codes did not have a higher average word count than participants from several other patterns (with fewer than four code types). See Supplemental Appendix S1 for more information.

### Research Question 3: Between-Group Differences in Content

Our final research question related to between-group differences in identity content at both the statement and pattern levels. For these analyses, we included participants for each question where they had demographic data available. Accordingly, the sample sizes differed across the tests we conducted.

In preparation for the analysis of between-group differences in codes, we examined between-group differences in the number of statements. There were differences based on racial identification,  $F(4, 393) = 3.21, p = .013$ . The average number of statements was slightly lower among participants who identified as Hispanic/Latino/Latina ( $M = 9.10$ ) or African American/Black/of African descent ( $M = 9.44$ ) than those who identified as multiracial ( $M = 9.67$ ), White/European American ( $M = 9.67$ ), or Asian/Asian American ( $M = 9.92$ ). Boys wrote fewer statements on average ( $M = 9.46$ ) than girls did ( $M = 9.73$ ),  $F(1, 407) = 5.56, p = .019$ . Regression analyses (with both linear and quadratic terms) showed that age did not predict the number of statements,  $R^2 = .01, p = .13$ .

We had intended to compare the number of types of codes (e.g., number of Personal codes) between groups. However, the between-group differences in the number of codable statements led us to rethink that strategy. Because some groups wrote (on average) fewer statements than others, they had fewer "chances" for codes to be applied, and thus fewer codes overall. Accordingly, between-group differences in numbers of codes were confounded with between-group differences in the number of statements, so they did not provide an unbiased estimate of the extent to which groups of participants may have included more of certain types of content (e.g., gender) than others.

To address this confounding, we used the average number of a certain type of code per statement (i.e., the quantity of a type of code within a statement set, divided by the number of statements in that set), which we refer to as codes per statement. To illustrate, consider two participants who both wrote two statements that were coded as Social Categories, but one wrote ten statements and the other wrote eight. Using the absolute number of Social Categories codes, the two participants have the same score. Dividing by the number of statements, however, shows that the participant with ten statements put less emphasis on Social

TABLE 3  
Description, Frequencies, and Examples of Patterns of Codes

All Main Content Codes (n = 24)	Personal, Social Categories, Relationships (n = 13)	Personal, Social Categories, Self-evaluation (n = 19)	Personal, Social Categories (n = 33)
(1) a person (2) a girl (3) a 13-year-old (4) smart (5) a Hunger Games fan (6) a [school name] student (7) one of seven children (8) a good person (9) imaginative (10) a writer	(1) a dancer (2) a stronger, independent woman (3) an animal lover (4) a student (5) a feminist (6) a little sister (7) a daughter (8) a person to cry on (9) a friend (10) strong	(1) quiet (2) unique (3) artistic (4) a day-dreamer (5) a test-taker (6) Chinese (7) proud of who I am (8) a good student (9) an American (10) a Christian	(1) an athlete (2) a boy (3) not a girl (4) funny (5) smart (6) tall (7) cool (8) popular (9) taking a survey right now (10) holding a pen
(1) fantastic at hockey (2) a good friend (3) nice to people (4) very athletic (5) very very strong (6) a great musician (7) a honor student (8) usually happy (9) funny (10) a fun person	(1) a good time (2) fun to be around (3) a friend (4) generous (5) respectful (6) understanding (7) sensitive to others (8) creative (9) loving (10) independent	(1) smart (2) fun to be around (3) happy (4) a great artist (5) a lovely person (6) nice to others (7) always smiling (8) very talkative (9) a good singer (10) a great dancer	(1) stubborn (2) sometimes annoying (3) usually witty (4) usually cocky (5) creatively (6) smart (7) keen (8) playful (9) myself (10) playful
Personal, Relationships, Self-evaluation (n = 66)	Personal, Relationships (n = 33)	Personal, Self-evaluation (n = 99)	Personal (n = 128)

Categories ( $2/10 = 0.20$  codes per statement) than the participant with eight statements ( $2/8 = 0.25$  codes per statement). The revised quantity of codes per statement did not differ by racial-ethnic identification,  $F(4, 393) = 1.99, p = .10$ , or gender,  $F(1, 407) = 0.10, p = .75$ .

**Content codes.** We identified few differences in content-related codes per statement based on participant characteristics. The number of Content codes per statement did not differ by racial identification,  $F(4, 393) = 1.21, p = .31$ , or gender,  $F(1, 407) = 0.29, p = .59$ . Regression analyses (with both linear and quadratic terms) showed it was not related to age,  $R^2 = .01, p = .30$ .

**Personal.** The average number of Personal codes per statement was slightly higher for girls ( $M = 0.94$ ) than for boys ( $M = 0.91$ ),  $F(1, 407) = 4.21, p = .04$ . There were no differences based on racial identification,  $F(4, 393) = 0.67, p = .61$ , or age,  $R^2 = .002, p = .67$ .

**Relationships.** The number of Relationships codes per statement did not differ based on gender,  $F(1, 407) = 2.35, p = .13$ , racial identification,  $F(4, 393) = 1.42, p = .23$ , or age,  $R^2 = .002, p = .63$ .

**Self-evaluation.** There were no differences in the average number of Self-evaluation codes per statement based on gender,  $F(1, 407) = 3.08, p = .08$ , racial identification,  $F(4, 393) = 0.65, p = .63$ , or age,  $R^2 = .01, p = .06$ .

**Social categories.** There were no differences based on gender,  $F(1, 407) = 0.67, p = .42$ , or racial identification,  $F(4, 393) = 1.30, p = .27$ . The number of Social Category codes per statement was, however, related to age,  $R^2 = .05, p < .001$ . The linear term for age was not statistically significant (unstandardized =  $-0.01$ , standardized =  $-0.19, p = .25$ ), but the quadratic term was (unstandardized =  $0.003$ , standardized =  $0.40, p = .017$ ). Accordingly, the predicted Social Category codes per statement were 0.03 for 11-year-olds, 0.023 for 12-year-olds, 0.022 for 13-year-olds, 0.027 for 14-year-olds, 0.038 for 15-year-olds, 0.055 for 16-year-olds, and 0.078 for 17-year-olds.

We next explored three specific Social Categories subcodes; we tested whether participants who (in the demographic section of the survey) identified as being part of a group marginalized in the United States (here, girls and youth of color) were more likely to include these group memberships in their responses. Girls were equally as likely as boys to include gender in their statements,  $\chi^2(1, n = 408) = 0.33, p = .57$  (the sample size of participants choosing an option other than boy or girl

was too small for follow-up analyses). However, the inclusion of racial group within a set of statements differed based on self-reported racial identification,  $\chi^2(4, n = 397) = 12.30, p = .015$ . A higher percentage of adolescents who reported that they identified as Black/African American/of African descent, Asian/Asian American, or multiracial had at least one statement coded as racial identity (11.29%, 5.40%, and 8.70%, respectively) compared with adolescents who chose White/European American (1.95%) or Hispanic/Latino/Latina (2.08%). The inclusion of an aspect of ethnic identity in the statements also differed based on racial identification,  $\chi^2(4, n = 397) = 18.81, p = .001$ . Adolescents who identified as Black/African American/of African descent, Asian/Asian American, or Hispanic/Latino/Latina were more likely to have statements with ethnic identity information (14.52%, 13.51%, and 14.6%, respectively) than adolescents who chose White/European American (2.92%) or multiracial (2.18%).

**Follow-up analyses of age differences.** Because girls and racial-ethnic minority participants were older on average, we conducted follow-up analyses for age separately by gender and in the five largest racial/ethnic groups (full results available from the first author upon request). In all but one case, the results were not substantively different. The only difference was related to racial-ethnic identification and Social Categories codes per statement. The overall pattern of more Social Categories codes per statement (described above) only held for participants who identified as White/European American (linear age term: unstandardized =  $-0.02$ , standardized =  $-0.41, p = .06$ ; quadratic age term: unstandardized =  $.005$ , standardized =  $0.53, p = .02$ ) or Asian/Asian American (linear: unstandardized =  $-0.22$ , standardized =  $-.78, p = .16$ ; quadratic: unstandardized =  $.01$ , standardized =  $1.15, p = .04$ ). For all other racial-ethnic groups, age did not predict Social Categories codes per statement.

**Structure codes.** The distribution of Structure codes per statement was skewed and leptokurtic, as nearly half of participants ( $n = 193; 46.5%$ ) had no Structure codes. Accordingly, we tested for between-group differences in the presence of at least one Structure code. The proportions of participants with and without Structure codes did not differ based on gender,  $\chi^2(1, n = 409) = 0.55, p = .46$ , or racial identification,  $\chi^2(4, n = 398) = 7.30, p = .12$ , and the average age was similar

between participants with a Structure code ( $M = 13.69$ ) and without ( $M = 13.49$ ),  $t(400) = 1.04$ ,  $p = .30$ .

**Patterns of codes.** Pattern was not related to gender,  $\chi^2(1, n = 408) = 8.46$ ,  $p = .29$ . For the relationship between pattern and racial identification, half of the cells had expected counts less than five (due to small numbers of both some patterns and some racial identification groups). Accordingly, we did not conduct further testing of that relationship. A one-way analysis of variance showed that there were differences in average age by pattern group,  $F(7, 401) = 11.65$ ,  $p = .002$ . Follow-up comparisons showed that the average ages differed significantly between the Personal and Social categories pattern ( $M = 14.56$ ;  $SD = 1.63$ ) and two other patterns: Personal, Relationships, Self-Evaluation ( $M = 12.97$ ;  $SD = 1.70$ ), and Personal, Self-Evaluation ( $M = 13.26$ ,  $SD = 1.81$ ).

**Word count.** The total number of words used by a participant for their set of statements ranged from 1 (i.e., a participant who wrote one statement of a single word,  $n = 1$ ) to 152 ( $n = 1$ ). The average words *per* statement ranged from 1 ( $n = 69$ ; 16.6%) to 15.20 ( $n = 1$ ) and were substantially positively skewed: 392 participants (95.2%) had an average word count per statement of 5 words or fewer. There were no between-group differences in average word count (see Appendix S1).

## DISCUSSION

The purpose of this analysis was to explore salient identity content reported by adolescents (specifically in the northeastern United States). We undertook this analysis because content is important but neglected in identity development literature (Gallagher et al., 2017). To elicit aspects of identity content, we used a shorter version of the TST (Kuhn & McPartland, 1954). We categorized the types of responses, identified patterns in content types, and investigated between-group differences in types and patterns.

### Research Question 1: Types and Prevalence of Content

Given our inclusive perspective about identity content, we coded the vast majority of adolescents' responses as being potentially or likely related to content, with only a small percentage of statements

being coded as related to Temporary States. Our inductive content analysis of the TST responses replicated some information about content obtained with other approaches that are more implicit and researcher-driven, but the findings also extend prior research in important ways.

**Replications of prior research.** Using the more open-ended format of the TST, our results replicate prior research with respect to the presence of content in the eight ideological and interpersonal domains traditionally used in the identity status literature (e.g., Balistreri et al., 1995). However, similar to studies where participants could adapt their responses to any domain (e.g., McLean, Syed, & Shucard, 2016), we found that the salience of these domains varied considerably. In addition, the overall pattern of results (with respect to which domains were identified more frequently) is consistent with the findings of McLean, Syed, and Shucard (2016), despite the differences in data collection methods between our study (using the TST) and theirs (using narrative prompts).

Within the interpersonal domain, references to friends and family were common, whereas romantic relationships were rarely mentioned. The higher prevalence of references to friends and family may be indicative of the emphasis placed on these roles in adolescence relative to young adulthood. Adolescents are often engaged primarily in exploration related to romantic relationships (e.g., Meier & Allen, 2009), and the short-term and temporary nature of these relationships may mean that they are not yet incorporated into adolescents' sense of self.

Aspects of ideology as previously defined in the identity status literature (religion, occupation, politics, values) were represented, although much less frequently than interpersonal aspects. Religious affiliation was explicitly mentioned only 12 times, and a spiritual or religious relationship (e.g., with God) was included in 17 statements. On the one hand, this low frequency is not surprising, given that American adolescents are not, on average, highly religious (King & Boyatzis, 2015). On the other hand, our sample was drawn primarily from religiously affiliated schools, where students may be more religious. From that perspective, the lower frequency is surprising; however, religiosity may not have stood out to participants because they attended school with others with a similar level of religiosity.

Direct references to occupation were rarely coded, which is not surprising given the age of

participants; a higher percentage might be expected beginning in young adulthood, when there is more attention to career choice (e.g., Roisman, Masten, Coatsworth, & Tellegen, 2004). However, aspects of *studenthood* were mentioned 96 times. Being a student is, in a sense, the occupation of adolescence. Schools play an important role in the lives of adolescents (Lannegrand-Willems & Bosma, 2006), and the implications of school for self-definition are established early (e.g., Meeus & Dekovic, 1995). These data were also collected at school, which may increase its salience for some participants.

Political affiliation was also rarely explicitly mentioned, which is also age-appropriate. Although adolescents are developing many aspects of their identity that are relevant to later political participation (e.g., values), they may not commit to an explicit party affiliation until they are eligible to vote. Research from the identity status tradition, indeed, has found that most adolescents were classified as “diffused” with regard to political identity (e.g., Waterman, 1985). Our results provide more support for the idea that political party identification is not a particularly salient component of identity for most adolescents between the ages of 11 and 17. However, linking TST responses in adolescence to data from adulthood could illuminate important identity-related precursors to later political affiliation and behaviors.

Finally, direct references to values (abstract principles) were coded only 12 times. This low frequency may have occurred for several reasons. As abstract principles, the influence of values on the self may not be consciously registered (Hitlin, 2011), and accordingly, values may be less accessible when contemplating the “Who am I?” question. Articulating the abstract nature of values may also pose a cognitive challenge for younger adolescents (e.g., Keating, 2013). In addition, although the construction of a personal moral system is well underway by adolescence, these frameworks are not often fully assimilated into the sense of self until adulthood (e.g., Hardy & Carlo, 2011). Finally, the nature of the TST prompt may have discouraged participants from responding with values, as doing so would require a more complicated response (e.g., as one participant wrote, “I am a person who values diversity”).

Our results also replicate prior research related to the presence of content beyond the eight domains. Also consistent with McLean, Syed, and Shucard (2016), our results suggest that these eight domains are not the only ones salient to adolescents. Quite a few of the other domains we

identified overlapped with specialty topics in the identity literature (e.g., sports fan; Koch & Wann, 2016). For example, hobbies were mentioned 263 times, and being a “fan” of something (e.g., a person or a sports team) was present in 12 statements.

*Beyond prior research.* The correspondence between some of our results and prior research is encouraging; this overlap points to the convergent validity of the TST approach to gathering information about content compared with other approaches that have been used (e.g., narrative prompts; McLean, Syed, & Shucard, 2016). In other words, when participants are given the opportunity to provide their own answers to the “Who am I” question, some of what they say aligns with what researchers have already been asking about. However, our results go beyond these similarities in several ways.

In our view, an important contribution of this work is the identification of content elements that have not typically been included in identity development research. These include physical characteristics, emotions, and personality characteristics. The prominence of physical features (coded in 190 statements; more frequent than the Social Category of friendship) may reflect the dramatic physical changes that happen during adolescence, and the corresponding importance placed on physical attractiveness as a measure of self-worth (e.g., Brinthaup & Lipka, 2012). Emotions were also commonly included (178 statements). Although identity research has not often focused on emotions, the frequency of these responses raises questions about their role in adolescents’ identity. For example, do they view emotions (particularly ones they experience often) as similar to personality characteristics?

Another feature of many statements was the inclusion of self-evaluations. Other researchers have categorized TST responses (i.e., the statements themselves) as positive, negative, or neutral (Hihara et al., 2019). We did not, however, given that whether the valence of content likely depends on context and research questions. Our intent was to capture the types of responses generally and leave it to other researchers to categorize based on their interests. However, some statements contained the participants’ *own* evaluations of themselves, or of another element of content, as positive/negative /neutral (e.g., “I am a good friend” or “I am good at basketball”). Although self-evaluations are often included in the related concept of self-esteem, these results suggest that self-evaluations have close connections with

content. For example, the prevalence of positive self-evaluation may indicate that what adolescents think they are “good” at—their talents or abilities—plays an important role in their identity. Future research could focus on teasing apart these self-evaluations related to specific content from adolescents' broader sense of self-esteem.

Finally, with respect to content categories, the sheer number of responses involving personal characteristics (beyond beliefs, values, hobbies) merits attention. Personality characteristics (e.g., the Big Five) have largely been left out of conversations about identity content, at least from a developmental perspective (e.g., they are not included in Galliher et al.'s framework), although they are included as aspects of identity within other models (e.g., neo-socioanalytic model of personality; Roberts & Nickel, 2021). Researchers may find utility in making granular distinctions about whether personality descriptors “count” as identity content, but it is clear from our results that when asked to answer the fundamental identity question of “Who Am I?” personal characteristics come to mind for most adolescents. Accordingly, eliminating them from consideration of content is likely missing an important source of the sense of sameness and continuity that underlies adaptive identity development.

Beyond categories that we coded as content, an unexpected result was related to the structure of how participants wrote the statements. In preanalysis discussions, we had not anticipated that the structure of the statements would stand out. However, after we started coding, we noticed immediately that there were variations in how participants wrote the statements, such as using qualifiers or conditionals. These responses—such as specifying situations in which certain conditions or aspects of the self are present—may indicate an ability of the TST to capture important nuances in content. The variation in linguistic and grammatical structures that participants used highlights the complexity of content, while pointing to the need for measurement tools that can capture that nuance.

The variation we observed in the structure of adolescents' responses also supports the view of identity as a dynamic social construct, as opposed to a fixed set of characteristics, and as an emergent product (i.e., something that is constructed at least in part during the responding itself) rather than a preexisting source of linguistic practices (Bucholtz & Hall, 2010). With its open format and instructions to write statements as if participants were “saying” the answers to themselves, the TST allows

participants to respond in a conversational manner, which might best reflect how identities are constructed through verbal discourse in everyday life (e.g., Bamberg, De Fina, & Schiffrin, 2011). For example, certain verbal devices are used to emphasize self-differentiation; we observed some of these and captured them with the “comparative” Structure code. We also saw indications of how some participants may have been navigating constancy and change in their sense of self, as evidenced through statements that we coded as containing different verb tenses (e.g., “I am becoming comfortable in my own skin”). Future research should examine the role that these structural differences in content may play in development.

### Research Question 2: Patterns of Content

For our second research question, we explored whether content types co-occurred in patterns. Although it is widely acknowledged that individuals' identity consists of multiple types of content (e.g., Galliher et al., 2017), to our knowledge this analysis is the first to directly examine different ways in which contents occur, especially within the sentence-completion format of the TST.

Using the presence of types of Content codes, we identified eight patterns. The most common pattern was only Personal codes, whereas the least common included all four main Content codes. The remainder included two or three types of Content codes. Some of the patterns were more likely to include Structure codes—primarily (but not only) patterns with more types of content codes. Accordingly, we speculate that both patterns and Structure codes may be indications of identity content complexity. Future research could investigate the predictors and consequences of these patterns.

In addition, the method that we used is only one of the many potential ways to look at patterns of identity content in the TST. Even within our coding scheme, future research could consider using more information from the subcodes; for example, within the Personal codes only pattern, many subtypes of Personal content were included, so patterns could be derived from only Personal content codes. Future research could also employ model-based pattern analyses (e.g., latent class analyses).

### Research Question 3: Between-Group Differences in Content and Patterns

Our final research question related to between-group differences in codes and patterns. With

respect to the salience of types of content broadly speaking, we identified some differences related to Social Categories that are consistent with prior literature regarding adolescent development.

First, age was related to the use of Social Category codes in general, such that older adolescents used a larger number of Social Category codes per statement. This overall finding is consistent with a growing importance of social categories as adolescents gain the cognitive abilities to engage in more abstract reasoning about their social group memberships and the meanings that these memberships have for them. However, we view these results with caution for several reasons. Age only predicted a small amount of variation in Social Categories codes per statement. Furthermore, follow-up analyses showed that the pattern only held (based on statistical significance of coefficients) for White/European American and Asian/Asian American adolescents. These were the largest racial-ethnic groups, and the other three groups were somewhat small. Accordingly, we recommend that these analyses be replicated in larger samples with sufficient representation of all racial-ethnic groups; in addition, these findings were based on age comparisons within cross-sectional data, so longitudinal research is needed.

Our results also provide support for the idea that adolescents who are members of marginalized groups in the United States experience these social positions as more salient (e.g., Meca et al., 2015), and therefore are more likely to include these aspects in their responses. Adolescents who reported identifying as Black/African American/of African descent, Asian/Asian American, or multiracial were more likely to have a statement coded with the Racial Identity subcode than adolescents who identified as White/European American or Hispanic/Latino/Latina. In turn, adolescents who chose Black/African American/of African descent, Asian/Asian American, or Hispanic/Latino/Latina were more likely to include statements that referenced their ethnic identity. These patterns were observed despite there being no difference based on racial-ethnic identification in the use of Social Categories codes more generally; in other words, adolescents from all racial-ethnic groups were equally likely to include information in their statements related to Social Categories generally, but adolescents from marginalized racial or ethnic groups were more likely to include *specific* Social Categories related to race or ethnicity. These results are consistent with other findings that people of color rate their racial or ethnic group identifications

as more salient than White people do (e.g., Bombay et al., 2010).

In contrast to the pattern of results regarding racial and ethnic identification, we did not observe differences based on gender identification; boys were similarly likely as girls to include gender in their statements. In this sample, the subgroup of adolescents who chose the option of “another gender identification” on the survey was too small for follow-up analyses. In addition, our survey did not include a question about sexual orientation (and sexual orientation was mentioned in only seven statements). Investigating whether sexual and gender minority youth include those aspects of their identities in their responses to the TST is an important question for future research.

### The (Minimal) Role of Word Count

At the suggestion of a reviewer, we also considered the role of word count in TST responses. Overall, the most consistent finding was unsurprising: Statements with more words received (on average) more codes. The number of meanings that can be conveyed by a single word is limited. It is likely challenging to produce single-word responses that include multiple elements, especially within the time-limited nature of survey data collection. Including more words permits the inclusion of more information and, by extension, more things that can be coded. However, our analyses suggest that this is where the role of word count likely ends (i.e., with the *amount* of codable information). At the statement level, word count was not related to the type or co-occurrence of codes, and at the participant level, average word count did not differ based on gender, age, or racial-ethnic identification. Accordingly, researchers who are interested in the number of codes given to specific statements or sets of statements should consider word count within their analyses, but such adjustments are less likely to be needed when researchers are interested in types of codes (rather than the absolute number).

### Limitations and Future Directions

Our results demonstrate the variety of content that is salient to US adolescents, but they should be interpreted considering the study's limitations. First, our findings are limited in generalizability. We summarized the types of responses that were given by a group of adolescents attending primarily religiously affiliated schools in the north-eastern United States. Responses are likely to

differ depending on where, when, and from whom the data were collected, and other specific features of the sample. Future research could investigate the situationally dependent nature of the TST responses more explicitly (e.g., through varying the collection method between written and verbal).

The a-contextual nature of the data collection also limited the information we had about participants' responses, as they were not asked to provide further details about their statements or explain whether or how their responses fit together. Future research using the TST could prompt participants to write about their statements to explicitly investigate identity configurations (Hammack et al., 2009; Schachter, 2004; Syed, 2010). For example, participants could be asked whether any of their content elements conflict, remain separate, or work together. Furthermore, identity processes could be linked to specific content through TST responses; for example, participants could be asked how committed they feel to specific elements or how much time they spend thinking about or reconsidering them. These types of data would enable stronger empirical connections between content and process.

The cross-sectional nature of our analyses is also a potential limitation. We identified between-person age-related differences suggestive of development; future research with a larger age range may illuminate more differences in identity content between people (such as age-related differences in valuing of various aspects of identity content) or other types of identity content that are more salient at different points in development. We examined responses to the TST at only one time point, which does not permit developmentally oriented conclusions. Future research could use our coding scheme to examine change and stability to provide information about within-person changes, which could illuminate the role of identity content in functioning across development.

## CONCLUSION

This research has provided important information about the identity content that is salient to adolescents, which has received limited research attention. Our coding of TST responses not only replicated prior research regarding domains of identity content but also provided novel findings regarding types of content (e.g., personality characteristics) and information about the way

adolescents wrote the statements (what we termed Structure codes). We also showed considerable between-participant variation in the types of content that they provided (i.e., patterns). These findings emphasize the importance of asking participants directly about their identity content and illustrate that the TST may be a useful way to do that in larger-scale studies. By incorporating more information about content, both on its own and in studies of processes, we can expand our understanding of how people become who are they are and view themselves to be.

## REFERENCES

- Balistreri, E., Busch-Rossnagel, N. A., & Geisinger, K. F. (1995). Development and preliminary validation of the Ego Identity Process Questionnaire. *Journal of Adolescence, 18*(2), 179–192.
- Bamberg, M., De Fina, A., & Schiffrin, D. (2011). Discourse and identity construction. In S. J. Schwartz, K. Luyckx, & V. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 177–199). New York, NY: Springer.
- Bayarl, P. S., Horton, K. E., & Jacobs, G. (2018). How do we describe our professional selves? Investigating collective identity configurations across professions. *Journal of Vocational Behavior, 107*, 168–181. <https://doi.org/10.1016/j.jvb.2018.04.006>
- Bombay, A., Matheson, K., & Anisman, H. (2010). Decomposing identity: Differential relationships between several aspects of ethnic identity and the negative effects of perceived discrimination among First Nations adults in Canada. *Cultural Diversity and Ethnic Minority Psychology, 16*(4), 507–516.
- Brinthaup, T. M., & R. P. Lipka (Eds.) (2012). *Understanding early adolescent self and identity: Applications and interventions*. Albany, NY: SUNY Press.
- Brittian, A. S., Umaña-Taylor, A. J., & Derlan, C. L. (2013). An examination of biracial college youths' family ethnic socialization, ethnic identity, and adjustment: Do self-identification labels and university context matter? *Cultural Diversity and Ethnic Minority Psychology, 19*(2), 177.
- Bucholtz, M., & Hall, K. (2010). Locating identity in language. In C. Llamas, & D. Watt (Eds.), *Language and identities* (pp. 18–28). Edinburgh, Scotland: Edinburgh University Press.
- Bussey, K. (2011). Gender identity development. In S. J. Schwartz, K. Luyckx, & V. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 603–628). New York, NY: Springer.
- Crocetti, E., Rubini, M., & Meeus, W. (2008). Capturing the dynamics of identity formation in various ethnic groups: Development and validation of a three-dimensional model. *Journal of Adolescence, 31*(2), 207–222. <https://doi.org/10.1016/j.adolescence.2007.09.002>

- Ferguson, Y. L., Ferguson, K. T., & Ferguson, G. M. (2017). I am AmeriBritSouthAfrican-Zambian: Multidimensional remote acculturation and well-being among urban Zambian adolescents. *International Journal of Psychology, 52*(1), 67–76. <https://doi.org/10.1002/ijop.12191>
- Galliher, R. V., McLean, K. C., & Syed, M. (2017). An integrated developmental model for studying identity development in context. *Developmental Psychology, 53*(11), 2011–2022.
- Goossens, L. (2001). Global versus domain-specific statuses in identity research: A comparison of two self-report measures. *Journal of Adolescence, 24*(6), 681–699. <https://doi.org/10.1006/jado.2001.0438>
- Hammack, P. L., Thompson, E. M., & Pilecki, A. (2009). Configurations of identity among sexual minority youth: Context, desire, and narrative. *Journal of Youth and Adolescence, 38*(7), 867–883.
- Hardy, S. A., & Carlo, G. (2011). Moral identity. In S. J. Schwartz, K. Luyckx, & V. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 495–513). New York: Springer.
- Harter, S. (2015). *The construction of the self: Developmental and sociocultural foundations*. New York, NY: Guilford.
- Hihara, S., Sugimura, K., Umemura, T., Iwasa, Y., & Syed, M. (2021). Positive and negative valences of identities: Longitudinal associations of identity content valences with adaptive and maladaptive functioning among Japanese young adults. *Development and Psychopathology, 1*–15. <https://doi.org/10.1017/S0954579421000043>
- Hihara, S., Umemura, T., & Sugimura, K. (2019). Considering the negatively formed identity: Relationships between negative identity and problematic psychosocial beliefs. *Journal of Adolescence, 70*, 24–32. <https://doi.org/10.1016/j.adolescence.2018.11.002>
- Hitlin, S. (2011). Values, personal identity, and the moral self. In S. J. Schwartz, K. Luyckx, & V. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 515–529). New York: Springer.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.
- Johnson, S. K., Buckingham, M. H., Morris, S. L., Suzuki, S., Weiner, M. B., Hershberg, R. M., ... Lerner, R. M. (2016). Adolescents' character role models: Exploring who young people look up to as examples of how to be a good person. *Research in Human Development, 13*, 126–141. <https://doi.org/10.1080/15427609.2016.1164552>
- Karás, D. (2015). *Identity domains in emerging adulthood – results of the qualitative research in Poland*. The Newsletter of European Association for Research on Adolescence, May 2014, p. 11–14.
- Keating, D. P. (2013). Cognitive and brain development. In R. M. Lerner, & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 4–84). New York, NY: Wiley.
- King, P. E., & Boyatzis, C. J. (2015). Religious and spiritual development. In R. M. Lerner, & M. E. Lamb (Eds.), *Handbook of child psychology and developmental science, Vol. 3. Socioemotional processes* (7th ed., pp. 975–1021). New York, NY: Wiley.
- Koch, K., & Wann, D. L. (2016). Team identification and sport fandom: Gender differences in relationship-based and recognition-based perceived antecedents. *Journal of Sport Behavior, 39*(3), 278–300.
- Kuhn, M. H., & McPartland, T. S. (1954). An empirical investigation of self-attitudes. *American Sociological Review, 19*(1), 68–76. <https://doi.org/10.4324/9781315129945-16>
- Lannegrand-Willems, L., & Bosma, H. A. (2006). Identity development-in-context: The school as an important context for identity development. *Identity, 6*(1), 85–113.
- Lerner, J. V., Wong, C. C., Weiner, M. B., & Johnson, S. K. (2021). Profiles of adolescents' character attributes: Associations with intentional self-regulation and character role model relationships. *Journal of Moral Education, 50*(3), 293–316. <https://doi.org/10.1080/03057240.2020.1755242>
- Lumma, A. L., Böckler, A., Vrticka, P., & Singer, T. (2017). Who am I? Differential effects of three contemplative mental trainings on emotional word use in self-descriptions. *Self and Identity, 16*(5), 607–628. <https://doi.org/10.1080/15298868.2017.1294107>
- Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology, 3*(5), 551–558. <https://doi.org/10.1037/h0023281>
- Martin, E. M., & Horn, T. S. (2013). The role of athletic identity and passion in predicting burnout in adolescent female athletes. *The Sport Psychologist, 27*, 338–348. <https://doi.org/10.1123/tsp.27.4.338>
- McLean, K. C., & Syed, M. (2015). Personal, master, and alternative narratives: An integrative framework for understanding identity development in context. *Human Development, 58*, 318–349.
- McLean, K. C., Syed, M., & Shucard, H. (2016). Bringing identity content to the fore: Links to identity development processes. *Emerging Adulthood, 4*, 356–364. <https://doi.org/10.1177/2167696815626820>
- McLean, K. C., Syed, M., Yoder, A., & Greenhoot, A. (2016). The role of domain content in understanding identity development processes. *Journal of Research on Adolescence, 26*, 60–75.
- McNaughton, S. M., & Billot, J. (2016). Negotiating academic teacher identity shifts during higher education contextual change. *Teaching in Higher Education, 21*, 644–658.
- Meca, A., Ritchie, R. A., Beyers, W., Schwartz, S. J., Picariello, S., Zamboanga, B. L., ... Benitez, C. G. (2015). Identity centrality and psychosocial functioning: A person-centered approach. *Emerging Adulthood, 3*(5), 327–339. <https://doi.org/10.1177/2167696815593183>
- Meeus, W., & Dekovic, M. (1995). Identity development, parental and peer support in adolescence: Results of a national Dutch survey. *Adolescence, 30*(120), 931–945.
- Meier, A., & Allen, G. (2009). Romantic relationships from adolescence to young adulthood: Evidence from

- the National Longitudinal Study of Adolescent Health. *The Sociological Quarterly*, 50(2), 308–335.
- Naudé, L. (2020). Being me: Content and context in South African adolescents' identity development. *Current Psychology*, <https://doi.org/10.1007/s12144-020-00737-w>
- Roberts, B. W., & Nickel, L. B. (2021). Personality development across the life course: A neo-socioanalytic perspective. In O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (pp. 259–283). London: The Guilford Press.
- Rohall, D. E. (2018, August). *Assessing private and public selves utilizing the TST and the TST-short*. Paper presented at the conference of Interpretative Sociology, Philadelphia, PA.
- Roisman, G. I., Masten, A. S., Coatsworth, J. D., & Tellegen, A. (2004). Salient and emerging developmental tasks in the transition to adulthood. *Child Development*, 75(1), 123–133.
- Rosenfeld, D. L., & Burrow, A. L. (2017). Vegetarian on purpose: Understanding the motivations of plant-based dieters. *Appetite*, 116, 456–463. <https://doi.org/10.1016/j.appet.2017.05.039>
- Schachter, E. P. (2004). Identity configurations: A new perspective on identity formation in contemporary society. *Journal of Personality*, 72(1), 167–200.
- Syed, M. (2010). Developing an integrated self: Academic and ethnic identities among ethnically diverse college students. *Developmental Psychology*, 46(6), 1590.
- Syed, M., & McLean, K. C. (2015). The future of identity development research: Reflections, tensions, and challenges. In K. C. McLean, & M. Syed (Eds.), *The oxford handbook of identity development*.
- Thoits, P. A. (2013). Volunteer identity salience, role enactment, and well-being: Comparisons of three salience constructs. *Social Psychology Quarterly*, 76(4), 373–398. <https://doi.org/10.1177/0190272513498397>
- Umaña-Taylor, A. J. (2011). Ethnic identity. In S. J. Schwartz, K. Luyckx, & V. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 791–809). New York, NY: Springer.
- Waterman, A. S. (1985). Identity in the context of adolescent psychology. *New Directions for Child and Adolescent Development*, 30, 5–24.
- Waterman, A. S. (2015). What does it mean to engage in identity exploration and to hold identity commitments? A methodological critique of multidimensional measures for the study of identity processes. *Identity: an International Journal of Theory and Research*, 15, 309–349.

### Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Appendix S1.** Full results of analyses including word count.