Aptitude complexes: Expanding our view of language aptitude

Freeborn, L.; Andringa, S.; Rispens, J.

DOI
10.1111/modl.70002

Publication date
2025

Document Version
Final published version

Published in
The Modern Language Journal

License
CC BY

DOI: 10.1111/modl.70002

The Modern Language Journal

ORIGINAL ARTICLE

Aptitude complexes: Expanding our view of language aptitude

Lani Freeborn[1,2] | **Sible Andringa**[2] | **Judith Rispens**[2]

[1]Department of Dutch Studies, University of Amsterdam, Amsterdam, The Netherlands

[2]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands

Correspondence
Lani Freeborn, Department of Dutch Studies, University of Amsterdam, Spuistraat 134, Amsterdam 1012 VB, The Netherlands.
Email:l.j.v.freeborn@uva.nl

Abstract
In second language (L2) research, aptitude is typically viewed as a multicomponential, cognitive construct. Yet we know that L2 learning is influenced by multiple learner individual differences (IDs) besides cognitive abilities and that these IDs interact. In this article, we consider the affordances and implications of working toward a broader theory of language aptitude. Inspired by the theory of aptitude complexes, we examine the joint functioning of cognitive, affective, conative, and contextual processes in L2 learning. We conducted a large, exploratory study of aptitude with 544 learners of L2 Dutch. We used exploratory graph analysis to estimate a partial correlation network of 25 variables, combined with a community detection algorithm to identify communities of closely related variables. Results showed five communities of aptitude variables: cognitive abilities, proficiency and use, motivation and effort, self-beliefs and emotion, and pro-social behaviors. Learners with higher levels of multilingualism and less external pressure to learn Dutch made more L2 gains. Results also showed that L2 use and attitudes toward the Dutch community had an effect on L2 gains, via proficiency level and length of residence. Our findings fit with the theoretical framework and highlight the importance of working toward a multivariate, integrative theory of language aptitude.

KEYWORDS
aptitude complexes, individual differences, language learning

It is widely known that second language (L2) learning is influenced by many individual difference (ID) variables from across cognitive, affective, conative, and contextual domains, and that these ID variables interact (Dörnyei, 2010; Ehrman & Oxford, 1995; Kormos, 2013; Serafini, 2017; Singleton, 2017; Thompson, 2013; Winke, 2013). Yet in the field of second language acquisition (SLA), aptitude has become synonymous with cognitive aptitude. Affective variables, such as personality and anxiety, and conative variables, such as motivation and willingness to communicate (WTC), are not typically viewed as components of aptitude. In this article, we challenge this traditional view and expand our conceptualization of aptitude beyond cognitive abilities. We report findings from a large, exploratory study of 544 learners of L2 Dutch. Inspired by Snow's (1987, 1989a, 1992) theory of aptitude complexes, we take a whole-person approach to the study of aptitude, considering the joint functioning of cognitive, affective, conative, and contextual processes in L2 learning. There is a lack of large-scale studies to investigate the combined influence of multiple ID variables in the same sample of learners, putting the role of cognitive abilities within the wider L2 learning context. We begin with a brief overview of research on cognitive aptitudes in SLA. We then focus our discussion around Snow's theory of aptitude complexes and SLA studies that have considered the influence of cognitive abilities on L2 learning in combination with ID variables from other domains.

BACKGROUND

Cognitive aptitudes in SLA research

The foundations of SLA aptitude research lie with the work of John Carroll in the development of the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959). The development of the MLAT was funded by the US government, who, after the Second World War, wanted to create a practical test for the army and air forces to determine who should receive intensive language training (Stansfield & Reed, 2004). At that time, cognitive psychology was becoming the dominant paradigm, and Carroll and Sapon took an information-processing approach to developing the MLAT (Sasaki, 2012; Stansfield & Reed, 2004). The test measures four cognitive abilities: phonetic coding ability, grammatical sensitivity, rote learning ability, and inductive language-learning ability. Despite the fact that the MLAT is not available to individual researchers (Language Learning and Testing Foundation, 2025) and has not changed since it was first published in 1959 (see Sasaki, 2012), it is still widely used in SLA research today (Li, 2016). It is important to remember that the MLAT was created to serve a practical function and was not developed within a theoretical framework. Carroll (1981) never limited the construct of aptitude to cognitive abilities, defining aptitude as "an individual's initial state of readiness and capacity for learning a foreign language, and probable facility in doing so given the presence of motivation and opportunity" (p. 86). When developing the MLAT, Carroll (1964) also studied the effects of noncognitive ID variables, including motivation, test anxiety, and attitudes toward foreign language study. In spite of this, conceptualizations of aptitude in SLA research have remained limited to cognitive aptitudes. Since the MLAT, a number of other language aptitude tests have also been developed, largely based on the MLAT and with a focus on cognitive abilities (e.g., Hi-LAB, Linck et al., 2013; CANAL-F, Grigorenko et al., 2000; LLAMA, Meara, 2005).

Snow's theory of aptitude complexes

Richard Snow was an educational psychologist who studied the role of aptitude in learning from instruction. In contrast to the traditional view, Snow (1992) took a holistic, multivariate approach to the study of aptitude, including any personal characteristic that influenced learning. In addition to cognitive abilities, Snow stressed the importance of including affective and conative processes. Taking a whole-person perspective, Snow focused on the joint functioning of cognitive, affective, and conative processes in learning from instruction (Snow, 1987). Much of his research was dedicated to the study of aptitude complexes, defined as "critical combinations of aptitude variables" (Snow, 1987, p. 19). Snow (1989a) also viewed aptitude as a dynamic construct that can be developed and influenced by students' beliefs about themselves as learners. Key to this conceptualization is how different aptitudes relate to the inner and outer environment. As Snow (1987) described, "an aptitude construct represents an interface between an inner environment—the substance of the artifact itself—and an outer environment in which it operates" (p. 12). For example, ID constructs such as motivation and anxiety belong to the inner environment of the L2 learner and also interact with the outer learning environment and L2 context.

To further develop this theory, Snow (1987) noted the importance of exploratory, correlational studies to analyze interactions between aptitudes (A–A) and aptitudes and outcome measures (A–O). As a way of "mapping the terrain," Snow (1989b) aimed to build and test networks of correlations among ID variables and learning outcomes. Building on Snow's work, Ackerman and Heggestad (1997) conducted a meta-analysis to inductively derive constellations of ID constructs, identifying four groups of variables that clustered together, which they named *trait complexes* (see Ackerman, 2003). These complexes contained both nonability and ability ID constructs, from across cognitive, affective, and conative domains. For example, the intellectual–cultural trait complex included crystallized intelligence, artistic interest, openness to experience, and intellectual engagement (Ackerman, 2003). Ackerman (2003) and Ackerman and Kanfer (2009) also conducted empirical studies to explore whether the same patterns of relationships were found between complexes and outcome measures in different educational and work settings.

Snow (1987) reasoned that by identifying interactions between aptitudes (A–A) and with learning outcomes (A–O), we could then extend our line of enquiry to analyze how aptitudes and learning outcomes interact with different instructional treatments (A–T–O). This idea refers to aptitude–treatment interaction (ATI), which Snow (1963) first developed for his doctoral thesis. For Snow, the ability to modify instructional treatment and experimentally manipulate aptitude–outcome relations was fundamental to aptitude theory. These ideas have been influential in the field of SLA; a number of ATI studies have investigated the relationships between cognitive aptitudes and L2 learning outcomes under different types of instructional treatment (see Yilmaz & Granena, 2019). Snow's work has also guided the development of some L2 aptitude theories. For example, Robinson's (2001, 2005) aptitude complexes/ability differentiation framework focuses on how IDs in cognitive aptitudes interact with different L2 learning conditions. Snow and his colleagues also initially focused on cognitive aptitudes—mainly out of practical necessity (Snow, 1992). It is important to note, however, that they recognized the need to later integrate ID constructs from other domains: "[We] knew we would have to cycle back to the study of 'aptitude complexes' that would add in constructs reflecting conative and affective processes in learning" (Snow, 1987, p. 17). Li (2024) has noted the confusion around the term "aptitude," and the differences between how aptitude is conceptualized in ATI research in the field of SLA and Snow's original framework. Li (2024) has suggested that "ATI" be changed to "ITI" (ID–treatment interaction) to better encompass the range of ID variables known to influence L2 learning and avoid further misinterpretations.

Multivariate approaches to aptitude in SLA research

While aspects of Snow's work have been explored in the field of SLA, there has remained a focus on cognitive abilities under the label of aptitude. Only a small number of studies have considered the joint functioning of ID variables across different domains in conjunction with cognitive abilities. In the next section, we review studies that examined a large number of variables in combination, across cognitive, affective, conative, and contextual domains of L2 learning. Note that in these studies, noncognitive ID variables were not conceptualized as components of aptitude; there is little work exploring the construct of aptitude in a more holistic way.

Some seminal studies have found that language-learning outcomes were predicted by multiple different ID variables from across different domains. For example, in a study entitled “Cognition Plus,” Ehrman and Oxford (1995) examined the learning gains of 855 students receiving training for 34 different languages. Besides cognitive abilities measured via the MLAT, Ehrman and Oxford included a range of other affective and conative ID constructs, such as learning styles, strategy use, motivation, anxiety, and personality. They also considered the role of demographic variables, including age, education level, and the number of languages previously learned. They conducted correlational analyses to explore the relationships between ID constructs and learners’ end-of-training speaking and reading proficiency scores. Results showed that the two outcome measures were most strongly related to cognitive abilities, followed by affective factors (motivation, self-confidence, and affective arousal) and personality variables. Demographic variables were also related to outcome measures, with an advantage for higher education level, younger age, and more previous language-learning experience.

In another seminal study called “Towards a Full Model of Second Language Learning,” Gardner et al. (1997) found that several ID constructs were interrelated and that L2 learning was predicted by multiple ID variables from different domains. With data from 102 learners of L2 French, Gardner et al. used exploratory factor analysis to determine the structural relationships among 29 ID variables. They then conducted correlations between aggregate (factor) scores and two outcome measures of L2 achievement. Overall, language anxiety, self-confidence, and can-do self-ratings of proficiency had the strongest associations with outcome measures. To further explore the relationships between the aggregated ID measures and L2 outcomes, Gardner et al. (1997) estimated a causal model of SLA using structural equation modeling (SEM). In their model, motivation, cognitive aptitude, and language-learning strategies were seen as direct predictors of language achievement. Motivation and cognitive aptitude explained around the same portion of the variance of L2 achievement. The model also showed relationships between predictor variables; language attitudes predicted motivation, while motivation predicted both self-confidence and language learning strategies. Gardner et al. (1997) emphasized that “existing research and theory make it clear that a number of variables relate to achievement in the L2, and it seems obvious that the variables do not operate independently of one another” (p. 347).

Other SLA studies have investigated the relationships between cognitive aptitudes and other ID variables, focusing on a small number of variables. For example, in a study of 494 intermediate and advanced learners of L2 English, Lee (2020) used SEM to model relationships between cognitive aptitude, motivation, vocabulary learning strategies, L2 use outside the classroom, and two outcome measures (vocabulary depth and breadth). Lee’s SEM model showed direct effects of cognitive aptitude and L2 use on outcome measures, and direct effects of motivation on strategy use and L2 use. Motivation had an indirect effect on outcome measures, mediated by L2 use. In another study, Winke (2013) investigated relationships between various ID constructs in a sample of 96 learners of L2 Chinese. Winke (2013) used SEM to model the associations between cognitive aptitude, strategy use, motivation, and L2 learning outcomes. Results showed that each of the three predictors explained a similar amount of variance in L2 learning outcomes. In addition, cognitive aptitude was negatively correlated with motivation. Winke (2013) reasoned that this may have been due to a compensatory relationship, whereby learners needed more motivation if they had low cognitive aptitude. Findings from many other studies point to a complex pattern of relationships between L2 learning outcomes and cognitive, affective, and conative aptitudes. For example, in a meta-analysis of 66 studies, Li (2016) found that cognitive aptitudes had negative associations with anxiety and weak associations with motivation. Other studies have found relationships between cognitive aptitudes and personality traits (Biedroń, 2011), attitudes toward L2 learning (Cochran et al., 2010), and first language (L1) abilities (Sparks et al., 2011, 2012).

Language aptitude is also linked to the L2 learning context. The relationship between L2 learning and cognitive, affective, and conative IDs has been found to vary across different proficiency levels (Artieda & Muñoz, 2016; Dunn & Iwaniec, 2022; Serafini, 2017; Serafini & Sanz, 2016). For example, Serafini (2017) conducted a longitudinal study with 87 learners of L2 Spanish at three proficiency levels. Cognitive aptitudes (two working memory [WM] tasks) were measured at the start of the semester, different motivational constructs were measured at the beginning and end of the semester, and L2 grammatical development was measured throughout the semester. Serafini used correlational analyses and scatterplots to visualize relationships among variables over time. Results indicated different patterns of relationships between variables for proficiency levels. For example, for beginner learners, phonological WM was positively related to anxiety, but for intermediate learners, phonological WM was related to effort to learn and attitudes toward the L2 context. Studies have also shown that aptitude is related to the level of multilingualism and previous language-learning experience. Having previously gone through the process of learning an L2, multilinguals tend to develop greater metalinguistic awareness, learning strategies, and a broader linguistic repertoire (Cenoz, 2013). Sáfár and Kormos (2008) conducted a longitudinal study of 61 learners of L2 English; 21 learners were at a regular secondary school, while 40 learners were at a bilingual school and received intensive instruction. Sáfár and Kormos compared learners' cognitive aptitude scores at the beginning and end of the academic year. Results showed that while both groups made gains, the group that received intensive instruction improved their aptitude scores significantly more. Other research has shown that even a small amount of prior L2 learning has a positive impact on cognitive aptitudes (Thompson, 2013). Level of multilingualism has also been found to interact with levels of anxiety and enjoyment (Botes et al., 2020; Thompson & Lee, 2012). These interactions have led researchers to question the interface between language aptitude and multilingualism: Does high aptitude encourage multilingualism, or does the experience of learning a language increase one's aptitude? (Singleton, 2017; Thompson, 2013).

Reconsidering our view of language aptitude

As illustrated from previous SLA studies described above, it is widely accepted that L2 learning is related to many other ID variables besides cognitive aptitudes. We also know that there are many interactions between cognitive abilities and ID variables from affective, conative, and contextual domains. In addition, we know that the role of cognitive abilities in L2 learning is partly dependent on learners' previous language-learning experience and the L2 context. In order to fully understand the role of cognitive abilities in learning, we need to take a whole-person approach, considering the joint functioning of ID variables from different domains. Other SLA scholars have called for a more integrative, broader conceptualization of aptitude, to include affective and conative characteristics as well as demographic variables (Dörnyei, 2010; Kormos, 2013; Serafini, 2017; Singleton, 2017; Thompson, 2013; Wen, 2021; Winke, 2013). Compared to the traditional, cognitive view of aptitude, a whole-person approach makes more sense not only from a theoretical perspective but also from a pedagogical perspective. As others have previously noted, although measures of cognitive aptitude tend to predict L2 learning, language learning would likely not start in the first place without conative IDs such as motivation (Dörnyei, 2005; Graham, 2021). We must also consider the relevance of language aptitude research to teachers and learners; while cognitive aptitudes are important, we also know that many other IDs influence learning. For example, in many educational contexts, having knowledge of L2 learners' memory capacity or phonetic coding ability is not particularly useful to learners or teachers. Weighing the importance of motivation and cognitive aptitudes, Graham (2021) has argued that "motivation wins hands-down, as something that is amenable to change, and hence something worth paying attention to" (p. 118). Limiting our notion of language aptitude to a set of stable cognitive abilities is also likely to impact learners' self-efficacy (Graham, 2021). Taking a more holistic view of aptitude allows for a fuller appreciation of the diverse range of abilities, emotions, and behaviors that contribute to L2 learning success. Besides the seminal studies described above, there are relatively few large-scale studies that have investigated the combined influence of multiple ID variables on L2 learning. While we know a lot about the relationships between isolated ID constructs and L2 learning, we know very little about how they interact as a whole. We need more multivariate studies with larger sample sizes to put the role of cognitive aptitudes in context and consider the bigger picture.

The present study

We conducted an exploratory study of language aptitude with 544 learners of L2 Dutch. Our overall aim was to study language-learning aptitude from a broader perspective, considering the joint functioning of ID variables in L2 learning. Following Snow (1989b), we aimed to build correlational networks to analyze interactions between ID constructs (A–A) and ID constructs and L2 learning gains (A–O). We initially included 27 measures across cognitive, affective, and conative domains the L2 context and participant demographics, and a measure of L2 learning gains (see Online Supporting Information A for the full list of variables). Our decisions about which ID variables to include were informed first by Snow's theory of aptitude complexes; we aimed to include ID constructs representative enough as to not miss potentially important learner characteristics across different domains. Second, we considered empirical SLA research on the role of IDs in language learning and multivariate studies of language aptitude. The choice of ID constructs was also dependent on the context of L2 Dutch learners in the Netherlands and pragmatic constraints of the study design. For example, questionnaires and tasks needed to be suitable for participants from different L1 backgrounds, not too time-consuming, and appropriate for online data collection.

Participants were enrolled in different L2 Dutch courses across the Netherlands. We collected data at two time points: at the start and end of the language courses. Following Snow's (1989a) description of aptitudes as "initial states of persons that influence later developments" (p. 876), we measured all variables at the start of participants' Dutch courses (T1), including a C-test to measure initial Dutch language proficiency. At the end of their Dutch course (T2), participants took the C-test a second time to assess their L2 gains. We used exploratory graph analysis (EGA) to estimate a partial correlation network of ID variables, combined with a community detection algorithm to identify communities of closely related variables. We aimed to answer the following research questions:

RQ1. What communities of ID variables emerge in a sample of L2 Dutch learners?
RQ2. Through which variables are communities related?
RQ3. Which variables are related to L2 learning and to which communities do they belong?

METHOD

All experimental materials and statistical analyses can be found in Online Supporting Information.[1]

Participants

Given the number of variables we aimed to analyze in this study, we needed a large sample of L2 learners. We cast a wide net, opening the study to anyone enrolled in an instructed L2 Dutch course, at any proficiency level, living in the Netherlands, from any L1 background. As such, a certain level of heterogeneity was inherent to the study design. Most participants were recruited from private language schools and university language centers. These settings tend to have structured Dutch courses and use a course book to cover reading, writing, speaking, and listening skills, and vocabulary development. Beyond this, we have limited knowledge of participants' instructed contexts. While a large number of people (1,299) joined the study, not all completed their aptitude profiles. There was also considerable variation between the number of participants who completed specific tasks. For example, 1,000 participants completed the personality questionnaire, while 689 participants completed the T1 C-test. In addition, some participants completed all tasks in the aptitude profile except for the C-test, while a small number of participants completed only the C-test. Rather than having large differences in sample size for different variables, and for reasons of simplicity and interpretability, we decided to include only participants who completed all tasks at T1 in analyses, which resulted in a sample of 544 participants.

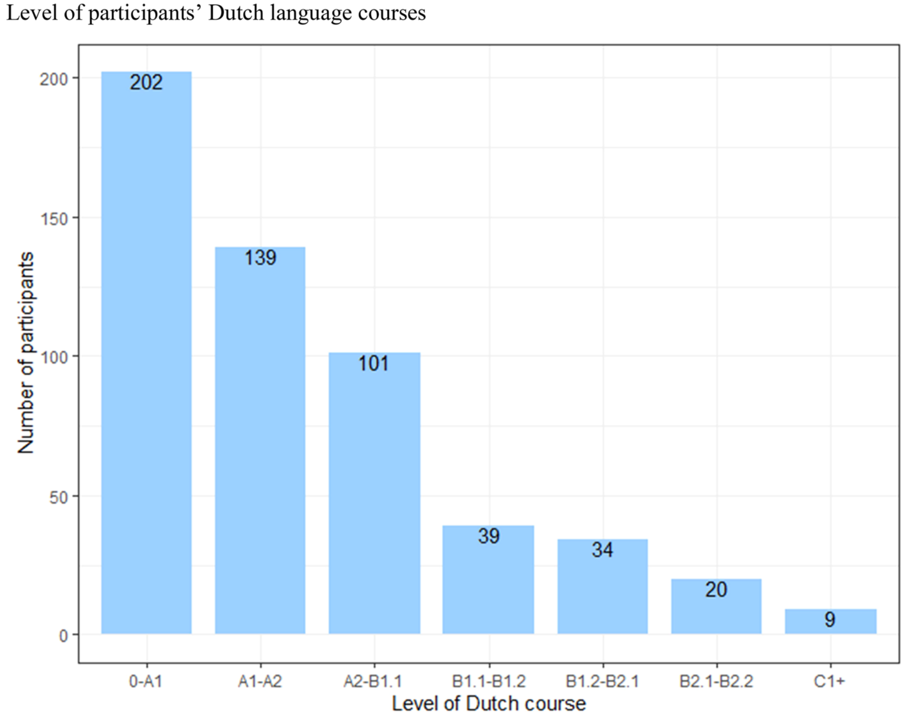

FIGURE 1 Level of participants' Dutch language courses. *Note*: Level reflects the Common European Framework of Reference for Languages. [Color figure can be viewed at wileyonlinelibrary.com]

Participants were 544 adults (66% female). On average, participants were 30 years old ($M = 30.50$, SD = 8.67), ranging from 18 to 75. They came from 61 different L1 backgrounds. The most common L1 background was English (n = 102), followed by Spanish (n = 58), Russian (n = 36), Turkish (n = 36), and German (n = 35). All participants reported at least an A2 English proficiency level in the Common European Framework of Reference for Languages. The majority of the participants in this study were highly educated; 90% of participants had earned, or were working toward, a university degree. Participants had been living in the Netherlands for varying amounts of time, with an average length of residence (LOR) of 2 years ($M = 26.02$ months, SD = 25.83). Some participants had recently arrived in the Netherlands (0 months), while the longest LOR was 14 years (172 months). Similarly, participants had varying amounts of previous experience learning Dutch. On average, participants had been learning Dutch for about 6 months ($M = 6.25$, SD = 10.95), ranging from 0 months (no previous instruction) to 108 months (9 years).

Dutch courses took place either online (48%) or face-to-face (52%), at private language schools and university language centers across the Netherlands. Figure 1 shows the levels of Dutch courses that participants were enrolled in. Most participants were enrolled in 0–A1 courses (37%), followed by A1–A2 courses (25%). We collected information about the start and end date of each participant's Dutch course and how many hours of instruction they received each week. There was a lot of variation in the duration and intensity of participants' Dutch language courses. On average, Dutch courses lasted for 8 weeks ($M = 8.57$, SD = 4.71), with 4 hours of instruction each week ($M = 4.17$, SD = 3.41).

TABLE 1 Descriptive statistics of the demographic and contextual variables ($N = 544$).

Variable	Mean (SD)	95% CI	Skewness (std. error)	Kurtosis (std. error)	Min	Max
Age (in years)	30.50 (8.67)	[29.77, 31.23]	1.67 (0.10)	4.57 (0.20)	18	75
LOR (in months)	26.02 (25.83)	[23.85, 28.19]	1.63 (0.10)	3.60 (0.20)	0	172
Amount of instruction[a]	29.42 (20.94)	[20.94, 27.66]	4.28 (0.10)	30.44 (0.20)	6	244
L2 use[b]	2.34 (2.73)	[2.11, 2.57]	2.07 (0.10)	5.63 (0.20)	0	16
Multilingualism[c]	7.65 (4.60)	[7.26, 8.04]	0.69 (0.10)	1.15 (0.20)	0	30

Abbreviations: CI, mean confidence interval; LOR, length of residence.
[a]Amount of instruction was calculated as the number of weeks multiplied by hours of instruction each week.
[b]L2 use was calculated as a sum score of use across four situations.
[c]Level of multilingualism was calculated by summing the proficiency level of each L2.

Measures

Contextual and demographic measures

We used a language background questionnaire to collect information about participant demographics and other contextual variables. All questionnaires used in the current study are included in Online Supporting Information C. As described in the previous section, participants provided information about their age, LOR, and the amount of instruction during their Dutch course. We created a sum score of the amount of instruction, multiplying the number of hours of instruction participants received each week, by the number of weeks the Dutch course lasted. We also asked participants about their L2 use. Participants were asked how often they used Dutch in four situations: at home, at work or school, with friends, and during free-time activities. Responses were scored on a 4-point Likert scale (0 = *never*, 1 = *1 hour a week*, 2 = *a few hours a week*, 3 = *a few hours for a few days a week*, and 4 = *multiple hours for multiple days a week*). For analyses, we created a sum score of L2 use for the four situations, for each participant. In addition, it was important to consider participants' previous language-learning experience. Participants were asked to list all the languages they knew, with a maximum entry of six languages. From this list, participants identified their L1(s) and self-reported their proficiency level of each listed L2[2] from six options (1 = *A1 low beginner*, 6 = *C2 high advanced*). To create a score for multilingualism, we summed the proficiency levels of each reported L2 for each participant.

We originally intended to add a measure of linguistic distance, as research in the Netherlands has shown that both L1 and L2 distance can influence L3 Dutch proficiency (Schepens et al., 2016). However, when following the procedures to measure lexical and morphological distance in Schepens et al. (2016), values were often missing for non-Indo-European languages, resulting in the loss of more than half the data.[3] While we recognize the role of linguistic distance in L2 learning,[4] we could not determine a meaningful measure of linguistic distance within the scope of this study that would not result in significant data loss, to create a continuous (or ordinal) variable for the main analyses.

Table 1 shows descriptive statistics for the contextual and demographic variables included in analyses.

Conative and affective measures

From Dörnyei's (2005) L2 motivational self-system framework, we included the ideal L2 self and the ought-to L2 self, with items adapted from Hiver and Al-Hoorie (2020). The ideal L2 self is a source of internal motivation related to learners' ability to envision themselves using Dutch in future situations, while the ought-to L2 self refers to different external pressures and obligations that learners feel to learn Dutch. From Gardner's (2004) Attitude/Motivation Test Battery (AMTB), we included integrative orientation, a more intrinsic source of motivation to learn Dutch to become closer to the Dutch community, and instrumental orientation as a more extrinsic source of motivation to learn Dutch for better career prospects. Often considered an outcome of motivated behavior (e.g., Hiver & Al-Hoorie, 2020; You et al., 2016), we included L2 WTC (MacIntyre et al., 1999), to assess how likely learners are to initiate communication in Dutch in different settings, using items adapted from Ryan (2009). We also measured learners' intended effort to learn Dutch, using items adapted from Hiver and Al-Hoorie (2020). From the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich & de Groot, 1990), we included the constructs of self-efficacy and self-regulation. Self-efficacy refers to students' beliefs about their learning abilities, while self-regulation refers to a combination of metacognitive strategies, cognitive strategies, and self-management during learning tasks (Pintrich & de Groot, 1990).

We also included several different affective variables. Using items adapted from the AMTB (Gardner, 2004), we measured learners' attitudes toward learning Dutch and attitudes toward the Dutch community. We assessed foreign language classroom anxiety, defined as "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning" (Horwitz et al., 1986, p. 128), using items from the short-form scale developed by Botes et al. (2022). We also included the 50-item IPIP (International Personality Item Pool) version of the Goldberg (1992) markers for the Big-Five factor structure to measure five personality traits: extraversion, agreeableness, conscientiousness, emotional stability, and intellect/imagination.

Table 2 has a list of the conative and affective variables with reliability analyses and descriptive statistics. The five personality traits were scored on a 5-point Likert scale following Goldberg (1992). All other variables were scored on a 6-point Likert scale (1 = *strongly disagree*, 6 = *strongly agree*). Cronbach's alpha and McDonald's omega showed good reliability estimates overall, with the lowest estimates for instrumentality ($\alpha = .69$, $\omega = .75$) and self-regulation ($\alpha = .73$, $\omega = .72$). We report the mean and standard deviation for each variable, including mean 95% confidence intervals. As shown in Table 2, participants had a particularly high level of integrativeness. Note that, following Snow (1992), we do not draw a sharp distinction between affective, conative, and cognitive processes. As Snow and Jackson (1997) reasoned, there is not a true partition between constructs from these domains, as they are interrelated: "All human behavior, especially including instructional learning and achievement, involves some mixture of all three aspects" (p. 1).

Cognitive measures

We included five different measures of cognitive abilities. WM refers to the cognitive ability of storing and processing information temporarily, in order to complete mental tasks (Wen, 2022). To measure simple WM, we used a forward digit span subtest from the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1997). In this task, participants were shown sequences of numbers and asked to reproduce the numbers in the same order. Participants' scores were the longest sequence of digits that they were able to recall (maximum 9). To measure complex WM (also known as executive WM), we used the short version of the operation span task developed by Oswald et al. (2015). This task contained both a storage component and a processing component; participants were shown sequences of numbers, then asked to judge whether some simple equations were true or false, and then asked to reproduce the

TABLE 2 Reliability analyses and descriptive statistics of conative and affective variables ($N = 544$).

Variable	*K*	Α	ω	Mean(SD)	95% CI	Skewness(std. error)	Kurtosis(std. error)
Ideal L2 self	5	.80	.78	4.58 (0.89)	[4.51, 4.66]	−0.52 (0.10)	−0.00 (0.21)
Ought-to L2 self	5	.83	.83	3.60 (1.21)	[3.50, 3.70]	−0.08 (0.10)	−0.74 (0.21)
Integrativeness	4	.80	.79	5.31 (0.70)	[5.25, 5.37]	−1.38 (0.10)	2.58 (0.21)
Instrumentality	4	.69	.75	4.35 (1.10)	[4.26, 4.45]	−0.51 (0.10)	−0.42 (0.21)
Willingness to communicate	8	.91	.90	3.82 (1.10)	[3.73, 3.91]	−0.22 (0.10)	−0.29 (0.21)
Intended effort	5	.82	.80	4.04 (0.95)	[3.95, 4.11]	−0.23 (0.10)	−0.19 (0.21)
Self-efficacy	9	.91	.90	4.54 (0.79)	[4.47, 4.60]	−0.68 (0.10)	1.55 (0.21)
Self-regulation	9	.73	.72	4.39 (0.67)	[4.33, 4.45]	−0.23 (0.10)	0.05 (0.21)
Attitudes to learning Dutch	4	.79	.83	4.71 (0.77)	[4.64, 4.77]	−0.59 (0.10)	0.49 (0.21)
Attitudes to the Dutch community	5	.81	.83	4.29 (0.79)	[4.22, 4.36]	−0.33 (0.10)	0.16 (0.21)
Anxiety	8	.89	.88	3.22 (1.03)	[3.13, 3.31]	0.27 (0.10)	−0.58 (0.21)
Extraversion	10	.90	.89	3.03 0.80	[2.95, 3.10]	0.11 (0.10)	−0.70 (0.21)
Agreeableness	10	.80	.78	3.78 (0.45)	[3.74, 3.82]	−0.54 (0.10)	0.14 (0.21)
Conscientiousness	10	.80	.81	3.75 (0.62)	[3.70, 3.81]	−0.46 (0.10)	−0.17 (0.21)
Emotional stability	10	.88	.87	2.90 (0.76)	[2.83, 2.96]	0.10 (0.10)	−0.45 (0.21)
Intellect/imagination	10	.76	.78	3.81 (0.57)	[3.76, 3.86]	−0.26 (0.10)	−0.14 (0.21)

Abbreviations: α, Cronbach's alpha; CI, mean confidence intervals; *k*, number of items; ω, McDonald's omega.

numbers. We followed the same procedures as Oswald et al. (2015) and used partial-credit scores, but after piloting the task online, we reduced the number of sets from 3–7 to 2–6.

We used the LLAMA D test (Meara, 2005) to measure sound recognition. The LLAMA D test assesses participants' ability to recognize repeated words in spoken language, which relies on phonetic coding ability and phonological WM, and has also been considered a measure of aptitude for implicit learning (Granena, 2013). In the training phase of the LLAMA D, participants heard a series of 10 words in an unfamiliar language. Then, in the testing phase, participants heard a second series of 40 words, containing both new words and words repeated from the first series. Participants were asked

TABLE 3 Descriptive statistics of cognitive measures ($N = 544$).

Variable (Task)	Mean (SD)	95% CI	Skewness (std. error)	Kurtosis (std. error)	Min	Max
Simple WM (Digit Span)	7.12 (1.38)	[7.00, 7.23]	−0.31 (0.10)	−0.52 (0.21)	3	9
Complex WM (Operation Span)	22.21 (6.43)	[21.67, 22.75]	−0.11 (0.10)	−0.36 (0.21)	3	36
Sound recognition (LLAMA D)	22.72 (7.53)	[22.09, 23.36]	−0.31 (0.10)	0.01 (0.21)	0	40
Nonverbal intelligence (Raven's Standard Progressive Matrices)	7.54 (1.36)	[7.42, 7.65]	−1.00 (0.10)	0.96 (0.21)	2	9

Abbreviations: CI, mean confidence intervals; WM, working memory.

TABLE 4 Mean reaction times and accuracy percentage of the flanker task in neutral, congruent, and incongruent trials ($N = 544$).

Trial type	Mean (SD)	95% CI
Flanker RTs (ms)		
Neutral	455.66 (87.55)	[448.30, 463.02]
Congruent	458.69 (103.69)	[449.98, 467.40]
Incongruent	497.64 (103.25)	[488.96, 506.31]
Flanker accuracy (%)		
Neutral	0.98 (0.04)	[0.97, 0.98]
Congruent	0.98 (0.05)	[0.98, 0.98]
Incongruent	0.92 (0.10)	[0.91, 0.93]

Abbreviations: CI, mean confidence intervals; RTs, reaction times.

to identify whether or not they heard each word before in the training phase. We followed procedures for the LLAMA v.3 as described by Rogers et al. (2023), awarding 1 point for each correct answer and deducting 1 point for each incorrect answer. To assess nonverbal intelligence, we used the abbreviated 9-item version of the Raven's Standard Progressive Matrices (RSPM) developed by Bilker et al. (2012). The RSPM measures abstract reasoning and problem-solving skills that are related to general intelligence (Raven & Raven, 2003). In this test, participants were asked to identify a missing part to complete a puzzle using the options given, with the puzzles increasing in difficulty. Participants were scored on the number of items they answered correctly. Descriptive statistics for the digit span, operation span, LLAMA D, and RSPM are reported in Table 3. Overall, these four variables showed good reliability; scores on the operation span task and LLAMA D were evenly distributed, but participants achieved relatively high scores on the digit span and RSPM.

The last cognitive measure we included was a flanker task adapted from Eriksen and Eriksen (1974) to assess inhibitory control, which refers to "the ability to deliberately inhibit dominant responses or competing representations" (Lehtonen et al., 2018, p. 401). In this task, participants were asked to focus their attention on an arrow and indicate whether the arrow was facing right or left by pressing a button as fast as possible. The task had congruent, incongruent, and neutral trials. In Table 4, we report the mean accuracy percentage and reaction time (RT) for each condition. Participants displayed slightly lower accuracy and slower RTs for the incongruent trials, indicating reliability. To calculate the flanker effect for each participant, we subtracted the mean RT for the congruent items from the mean RT for the incongruent items (using only trials with correct responses).

C-test

We used a C-test to assess participants' Dutch language proficiency (available in Online Supporting Information B). C-tests are based on the principle of reduced redundancy testing and belong to the same family as cloze tests, dictations, and elicited imitation tasks (Raatz & Klein-Braley, 2002). The C-test used in this study consisted of eight short texts of increasing difficulty level. The first sentence of each text was kept intact; then the second half of every second word was deleted. Participants were asked to complete the missing words. Following procedures in Raatz and Klein-Braley (2002), participants were given 5 minutes for each text and could choose to move to the next text when ready (by clicking "next"). Each text contained between 22 and 25 gaps, with a total of 188 gaps (items) across the eight texts. C-tests are often considered a proxy measure of general L2 proficiency because they require both receptive and productive skills (Eckes & Grosjahn, 2006), and rely on lexical, morphological, syntactical, and graphological knowledge at the sentence level, as well as cohesion and organization at the text level (Raatz & Klein-Braley, 2002). Many studies have shown that C-tests tend to correlate with criterion-test scores (for an overview, see McKay et al., 2021). C-tests have also been used in several longitudinal studies to assess language-learning gains over time (e.g., Kliesch & Pfenninger, 2021; Schnoor et al., 2023). They are also well suited for collecting data online and with a larger number of participants, as they can be scored automatically.

We used item response theory (IRT) to score the C-test, following other studies that have used this method (Lee-Ellis, 2009; Mozgalina & Ryshina-Pankova, 2015; Paxinou et al., 2021). By using an IRT (Rasch) model, we can consider both person ability and item difficulty (Paxinou et al., 2021). A correct answer depends on the interplay between the student's language ability and the difficulty of the item. By taking both person ability and item difficulty into account, we are able to calculate a more accurate measure of Dutch language proficiency. Participants took the same C-test twice, at the start (T1) and end (T2) of their Dutch language course. We used IRT to calculate an ability score (θ) for each learner at T1 and T2 with the dexter package (Maris et al., 2024) in *R* (R Core Team, 2023). We used the T1 ability score as a measure of the participants' proficiency level at the start of the Dutch course. To measure L2 gains, we subtracted the T1 ability score from the T2 ability score.

To assess the construct validity of the C-test, we looked at correlations between the T1 C-test (proficiency level), L2 gains, and other contextual variables. The correlation table is provided in Online Supporting Information D. Out of the 544 participants, 384 returned to take the second C-test after finishing their Dutch course. Thus, correlations between L2 gains and other variables were based on 384 participants. Proficiency level, as measured by the T1 C-test, was strongly correlated with the level of Dutch course participants were enrolled in ($r = .73$, $p < .001$) and the amount of time participants had been learning Dutch ($r = .54$, $p < .001$). L2 gains were negatively correlated with proficiency level ($r = -.45$, $p < .001$) and course level ($r = -.39$, $p < .001$). LOR showed weaker correlations with the other L2-related measures. Overall, these correlations suggested high construct validity of the C-test.

Procedure

Data collection took place from January 2022 to August 2023. To recruit participants, we contacted teachers, coordinators, and directors of private language schools and university language centers across the Netherlands and posted on social media platforms. The study was conducted in English, as we included participants from diverse L1 backgrounds. When recruiting participants, we targeted language schools and universities that we knew provided Dutch language courses for more highly educated learners, likely to have an intermediate level of English. The study was piloted with a few people with approximately an English proficiency level of A2–B1, to assess how well they could understand the task instructions and questionnaires, as well as to assess the functionality of the website. While all participants reported having at least an A2 English proficiency level, with online data collection and self-report measures, we cannot be certain whether English proficiency level influenced participants' understanding and/or responses, and we acknowledge this limitation. Schools informed their students about the study via email with a text that we provided, as well as distributing flyers. Some schools and teachers invited us to their classes to explain the study and invite students to join, for both online and in-person classes. Participants joined the study at the start of their Dutch language course. For shorter, intensive courses, the start referred to the first 1–2 days, while for longer, extensive courses, the start referred to the first 1–2 weeks. Students who joined the study later in their Dutch course (≥ one third) were excluded from analyses.

All data were collected online via a website designed for this study (www.aptitudeproject.org). Within the website, questionnaires were linked to Qualtrics (2022), while the cognitive tasks and C-test were linked to Experiment Designer (Vet, n.d.) When joining the study, participants provided informed consent and information about their Dutch language course. Upon joining the study (T1), participants were asked to complete their aptitude profile. Their profile consisted of nine tasks: four questionnaires, four cognitive tests, and the C-test. Participants could choose to do the tasks in any order and could complete their profile over a few days (they did not need to complete all tasks in one sitting). We combined multiple variables into the same task; for example, in the task we called "Memory," participants did both the digit span and operation span test. The task we called "Motivation" had 31 items and measured six variables (ideal L2 self, ought-to L2 self, integrativeness, instrumentality, intended effort, and WTC). It took approximately 1 hour and 20 minutes to complete the nine tasks in the aptitude profile. Access to the second C-test was blocked until the end of the Dutch course, when participants received an email reminder to take the C-test again.

To incentivize learners to participate, we gamified the study experience by creating an attractive, interactive website that updated upon completion of each task. After each task, participants received brief feedback on what the task measured, how it related to language learning, and their score. As an additional incentive for participants to return at T2, we had a prize draw whereby 1 in 10 participants who completed the T2 C-test received a €20 online voucher. Considering the time needed to complete the aptitude profile and the longitudinal nature of this study, there is likely a selection bias in our sample, with participants who are particularly interested in language learning.

Statistical analysis

Out of the 544 participants, 384 returned to take the second C-test. This meant that we could only look at relationships between L2 gains and other variables with a smaller sample of 384 participants, while relationships between all other variables were based on the larger sample of 544 participants. As a preliminary first step, we conducted Spearman bivariate correlations between variables. The full correlation table is available in Online Supporting Information D. We then conducted an EGA to estimate a partial correlation network of ID variables, grouping together closely related variables with a community detection algorithm. We conducted all analyses using *R* (R Core Team, 2023). In this article, we show only the final model and accompanying information, but we report all the steps described below in full in Online Supporting Information D.

EGA

EGA combines network analysis with a community detection algorithm to identify dimensional structures of multivariate data (Golino & Epskamp, 2017). Networks are typically estimated using the EBICglasso, which estimates a Gaussian graphical model (GGM; Epskamp & Fried, 2018). With the GGM, nodes represent variables and edges represent partial correlations between two variables (after conditioning on all variables). With EGA, variables that are more strongly correlated are grouped into communities in the network. While network models are statistically equivalent to factor models, they have different theoretical underpinnings. In factor models, variables are thought to correlate because they measure the same latent construct, whereas in network models, groups of variables that are strongly correlated are seen as an emergent system, whereby variables are causally coupled (Christensen & Golino, 2021b). In this sense, a network community refers to a dimension of a dataset, consisting of a group of closely related nodes. In a similar way that factor loadings help identify which variables belong to a latent factor, network loadings can be used to identify which variables belong to a community. As Christensen and Golino (2021b) explained, network loadings represent "each node's unique contribution to the emergence of a coherent *dimension* [emphasis in original] (or collection of related variables whose relations are not necessarily due to a common cause)" (p. 1564). Following the recommended workflow of EGA with the package EGAnet (Golino et al., 2020b), we first determined redundancies using unique variable analysis (UVA), then performed the EGA, and then assessed the stability of the EGA.

UVA

We first conducted a UVA to identify and remove local dependence in the data. For this, we used the full dataset ($N = 544$) of 27 ID variables measured at the start of the participants' Dutch course, excluding L2 gains (a list of all variables can be found in Online Supporting Information A). UVA estimates a network and uses weighted topological overlap to assess the extent to which two variables have similar partial correlations to other variables in the network (see Christensen et al., 2023). Variables that strongly overlap are seen to represent redundant components of the system, with an overlap statistic of $> .30$ indicating a large redundancy (Christensen et al., 2023). An advantage of the UVA procedure is that it uses regularization techniques, shrinking some partial correlation coefficients to zero. This results in a sparser network, with less likelihood of upward bias (see Christensen et al., 2023).

Our UVA showed a large redundancy (.38) between ought-to L2 self and instrumentality, which are both measures of extrinsic motivation. When two variables show redundancy, by default the UVA retains the variable with the lowest maximum weighted topological overlap to other variables, and removes the other variable (Golino & Christensen, 2024). This resulted in retaining ought-to L2 self and removing instrumentality. The UVA also showed moderate redundancy (.26) between the ideal L2 self and WTC. As we know these two constructs to be theoretically distinct, we retained both the ideal L2 self and WTC for subsequent analyses.

EGA and community detection

We began the EGA with 27 variables (the 26 remaining variables measured at T1 plus L2 gains). We estimated a GGM using the EBICglasso, using Spearman correlations due to the different scales used to measure the ID variables. We estimated the network with missing data as pairwise, which computed correlations for all available cases between two variables. This means that correlations between L2 gains and other ID variables were based on the smaller sample of 384, while correlations between all other ID variables were based on the larger sample of 544. To identify communities of related variables, we used the Louvain community detection algorithm. The Louvain algorithm identifies hierarchical structures in the network by iteratively exchanging variables between communities until the modularity no longer improves (Christensen et al., 2023). We used the total entropy fit index (TEFI) to assess the fit of the model, whereby lower TEFI values indicate a better solution (Golino et al., 2020a). In the first analysis with 27 variables, the TEFI was 24.79. Network loadings showed that three variables did not belong to any community: inhibitory control, amount of instruction, and level of multilingualism. Spearman bivariate correlations also showed that these variables had fewer and weaker associations with other variables (see Online Supporting Information D). We conducted three separate EGAs, removing each variable in turn and reassessing the TEFI value. Removing inhibitory control lowered the TEFI value to −24.28, and removing both inhibitory control and the amount of instruction lowered the TEFI value to −23.25. Removing the level of multilingualism slightly increased the TEFI value to −23.30, meaning that removing this variable did not improve model fit. The final model thus included 25 variables (24 ID variables measured at T1 and L2 gains). With the final model, we calculated network loadings to determine how strongly each node contributed to each community and to identify which variables displayed cross-loadings, potentially belonging to more than one community. Following Christensen and Golino (2021b), we considered a network loading of .15 as substantial, which is equivalent to a factor loading of .40.

Stability analyses

To assess the stability of the empirical EGA, we performed a parametric bootstrap EGA (Christensen & Golino, 2021b) using the Louvain community detection algorithm and the default of 500 iterations. We compared the empirical structure with the median bootstrap structure, using 95% confidence intervals to assess the stability of the median number of communities. Following the EGA workflow (Golino et al., 2020b), we used summary statistics to determine the structural consistency of the empirical EGA across bootstrap samples. We considered values greater than .70–.75 to reflect sufficient structural consistency (Christensen & Golino, 2021a). Last, to identify which variables accounted for higher or lower structural consistency, we looked at the stability of each variable in each community.

RESULTS

Figure 2 shows the results of the EGA. Our first RQ concerned which communities of variables emerged. The EGA resulted in five communities, which we named to best represent the variables contained, considering which variables had the highest network loadings. The network loadings for the EGA in Figure 2 are reported in Table 5.

Community 1 is best defined as *cognitive abilities*. This community comprised four cognitive variables and age. The two WM tasks (operation span and digit span) had the strongest network loadings, followed by sound recognition (LLAMA D) and age. Age was negatively associated with cognitive abilities, while the four cognitive abilities showed positive associations. Although nonverbal intelligence (RSPM) was positioned in Community 1, it did not display significant network loadings.

We named Community 2 as *proficiency and use*. This community contained five variables. Proficiency level, as measured via the C-test at the start of the study, showed the highest network loadings, followed by time living in the Netherlands, L2 gains, and L2 use. There were positive relationships between proficiency level, time living in the Netherlands (LOR), and L2 use. L2 gains were negatively related to proficiency level and time living in the Netherlands. Although attitudes toward the L2 community also joined this community, it did not display significant network loadings.

Community 3 has six variables and is best defined as *motivation and effort*. In order of the strength of network loadings, Community 3 consisted of ideal L2 self, intended effort, integrativeness, WTC, ought-to L2 self, and multilingualism. As seen in Figure 2, there are positive associations between the three motivational constructs and intended effort and WTC. Multilingualism was placed in Community 3 because it showed weak, negative associations with ought-to L2 self and intended effort. However, multilingualism did not display significant network loadings for any community.

We labeled Community 4 as *self-beliefs and emotion*. This community comprised six variables: Self-efficacy had the highest network loadings, followed by self-regulation, anxiety, attitudes toward L2 learning, conscientiousness, and emotional stability. Within this community, there were positive associations between self-regulation, self-efficacy, conscientiousness, and attitudes toward L2 learning. Anxiety was negatively associated with emotional stability and self-efficacy.

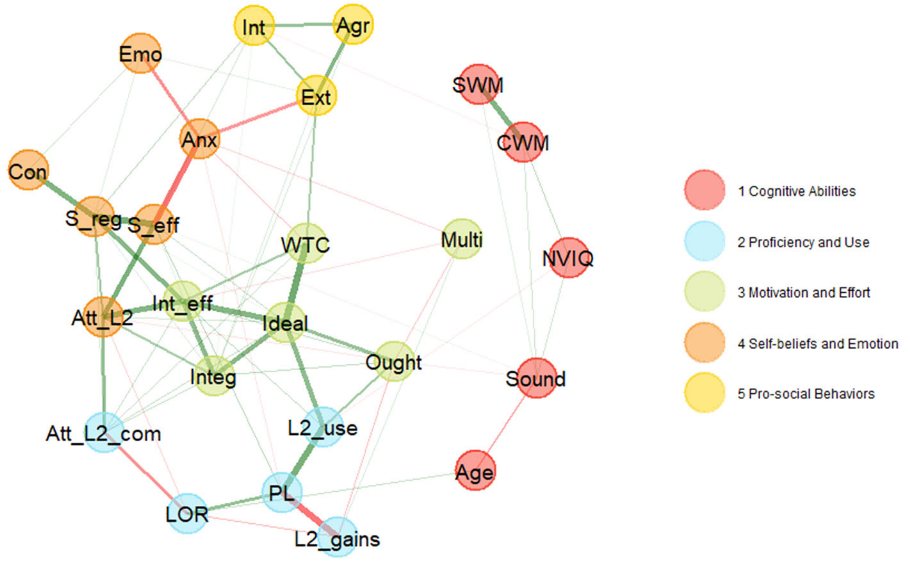

FIGURE 2 Exploratory graph analysis of 25 variables, resulting in five communities. CWM, complex working memory (operation span); SWM, simple working memory (digit span); Sound, sound recognition (LLAMA D); NVIQ, nonverbal intelligence (Raven's progressive matrices); PL, proficiency level; Att_L2_com, attitudes toward the Dutch community; LOR, length of residence; Ideal, ideal L2 self; Int_eff, intended effort; WTC, willingness to communicate; Ought, ought-to L2 self; Multi, level of multilingualism; S_eff, self-efficacy; S_reg, self-regulation; Att_L2, attitudes toward learning Dutch; Con, conscientiousness; Emo, emotional stability; Anx, language classroom anxiety; Agr, agreeableness; Ext, extraversion; Int, intellect/imagination. Edges between nodes represent partial correlations, with green and red edges showing positive and negative correlations, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

We named Community 5 as *pro-social behaviors*. This community contained three personality traits related to positive behaviors toward others: agreeableness, extraversion, and intellect/imagination. These three variables were positively associated.

With our second RQ, we aimed to determine through which variables the communities were related. The graph in Figure 2 shows that while the ID variables formed distinct communities, there were also many relationships between variables across different communities. Overall, cognitive abilities had fewer associations with variables from other communities, compared to the other four communities. Table 5 shows that several variables had substantive cross-loadings (≥ .15) across more than one community. L2 use had moderate network loadings on Communities 2 and 3. Intended effort and attitudes toward L2 learning both had moderate-to-high network loadings on Communities 3 and 4. Attitudes toward the Dutch community did not display significant network loadings for any community, but as shown in Figure 2, this variable was weakly related to six variables from three different communities. Overall, the graph in Figure 2 and the network loadings in Table 5 show that communities of variables were related mainly through L2 use, intended effort, and attitudes.

As we measured L2 use with a sum score across four different settings (at work or school, at home, with friends, when doing hobbies or free-time activities), we conducted additional analyses to explore whether there were any differential effects of L2 use across the four settings. We repeated the EGA four times (see Online Supporting Information D), with each of the L2 use settings. Results were overall very similar to the original model in Figure 2, but there were slight differences in the relationships between the L2 use contexts and motivational constructs. L2 use with friends and when doing hobbies was most strongly related to proficiency level and ideal L2 self, but was not related to ought-to L2 self. Slightly different relationships were found with L2 use at work or school and at home, as these items were most strongly related to ought-to L2 self, followed by proficiency level and ideal L2 self.

TABLE 5 Network loadings of the five communities.

Variable	Cognitive abilities	Proficiency and use	Motivation and effort	Self-beliefs and emotion	Pro-social behaviors
CWM	**.31**				
SWM	**.23**				
Sound	**.18**				
NVIQ					
Age	**−.10**				
PL		**.42**			
L2_use		**.15**	.17		
Att_L2_com					
L2 gains		**−.21**			
LOR		**−.23**			
Ideal		.13	**.43**		
Int_eff			**.27**	.21	
Integ			**.21**		
WTC			**.19**		
Ought			**.14**		
Multi					
S_eff				**.40**	
S_reg			.11	**.31**	
Att_L2		.11	.17	**.18**	
Con				**.13**	
Emo				**.11**	
Anx				**−.22**	−.12
Agr					**.30**
Ext					**.28**
Int					**.23**

Note: Showing standardized network loadings ≥ .10. To interpret network loadings: small = .15, moderate = .25, large = .35. A network loading of .35 is roughly equivalent to a factor loading of .70 (Christensen & Golino, 2021b).

Abbreviations: Agr, agreeableness; Anx, language classroom anxiety; Att_L2, attitudes toward learning Dutch; Att_L2_com, attitudes toward the Dutch community; Con, conscientiousness; CWM, complex working memory (operation span); Emo, emotional stability; Ext, extraversion; Ideal, ideal L2 self; Int, intellect/imagination; Int_eff, intended effort; LOR, length of residence; Multi, level of multilingualism; NVIQ, nonverbal intelligence (Raven's progressive matrices); Ought, ought-to L2 self; PL, proficiency level; S_eff, self-efficacy; S_reg, self-regulation; Sound, sound recognition (LLAMA D); SWM, simple working memory (digit span); WTC, willingness to communicate.

Our third RQ concerned the relationship between ID variables and L2 gains. From Figure 2, we see that L2 gains were related to four variables from Communities 2 and 3. In Community 2, proficiency level had the strongest association with L2 gains: Learners of lower proficiency levels made more gains compared to higher proficiency levels. Linked to proficiency level, time living in the Netherlands was also negatively associated with L2 gains. From Community 3, ought-to L2 self was negatively related to L2 gains: Participants who felt more obligated to learn Dutch made fewer gains. Also from Community 3, there was a positive relationship between L2 gains and the level of multilingualism.

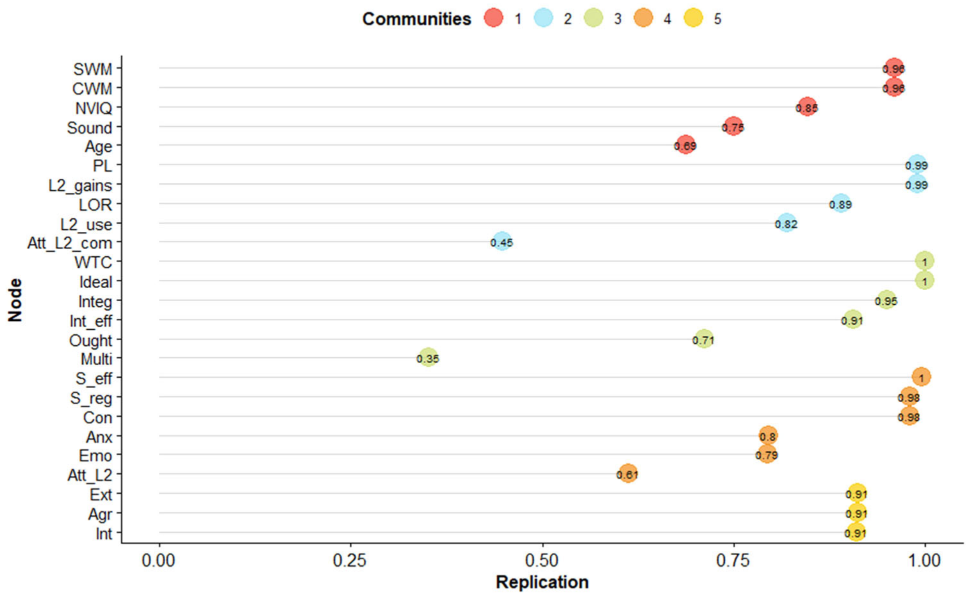

FIGURE 3 Results of the dimension stability analysis. CWM, complex working memory (operation span); SWM, simple working memory (digit span); Sound, sound recognition (LLAMA D); NVIQ, nonverbal intelligence (Raven's progressive matrices); PL, proficiency level; Att_L2_com, attitudes toward the Dutch community; LOR, length of residence; Ideal, ideal L2 self; Int_eff, intended effort; WTC, willingness to communicate; Ought, ought-to L2 self; Multi, level of multilingualism; S_eff, self-efficacy; S_reg, self-regulation; Att_L2, attitudes toward learning Dutch; Con, conscientiousness; Emo, emotional stability; Anx, language classroom anxiety; Agr, agreeableness; Ext, extraversion; Int, intellect/imagination. [Color figure can be viewed at wileyonlinelibrary.com]

Stability analyses

To assess the stability of the empirical EGA in Figure 2, we conducted a parametric bootstrap EGA. The full analyses and output are available in Online Supporting Information D. In the bootstrap EGA, the median number of communities was five, 95% CI [4.08, 5.92]. Out of the 500 iterations, 7% resulted in four communities, 77% resulted in five communities, and 14% resulted in six communities. However, the five communities generally showed low levels of structural consistency. Figure 3 shows the results of the dimension stability analysis, with the proportion of times that variables were assigned to the same community across the bootstrap replications. Community 1 had an overall consistency of 0.59, lower than the recommended level of 0.70 (Christensen & Golino, 2021a). This was partly explained by age, which appeared in 69% of bootstrap replications. Communities 2, 3, and 4 displayed low structural consistency of 0.32, 0.28 and 0.41, respectively. In Community 2, the low structural consistency was explained by attitudes toward the Dutch community, which appeared in 45% of replications. In Community 3, multilingualism had the lowest item stability, appearing in 35% of replications. In Community 4, attitudes toward L2 learning appeared in 61% of replications. Only Community 5, which consisted of three personality traits, showed high structural consistency (0.91).

Overall, the results of the stability analyses reflected the graphical positioning of each variable within the network in Figure 2, as well as the cross-loadings in Table 5. We can be fairly certain that there were five communities of ID variables in our data sample. The variables that appeared in fewer bootstrap replications tended to be the same variables that displayed network loadings for more than one community. L2 use, intended effort, and attitudes appeared to play a role in bridging the communities of proficiency and use, motivation and effort, and self-beliefs and emotion.

DISCUSSION

We conducted an EGA of 25 variables across cognitive, affective, conative, and contextual domains of L2 learning. Our first RQ concerned the relationships between ID variables and whether we could identify communities of closely related variables. Five communities of variables emerged from the data: cognitive abilities, proficiency and use, motivation and effort, self-beliefs and emotion, and pro-social behaviors. Each of these communities makes a lot of sense. The motivation and effort community, for example, showed close interrelations between types of motivation (ideal L2 self, ought-to L2 self, integrativeness) and motivated behavior (intended effort, WTC). SLA studies have shown similar associations between these variables, such as stronger associations between the ideal L2 self and intended effort, compared to the ought-to L2 self (see Al-Hoorie, 2018; Qi, 2022). Community 4, self-beliefs and emotion, also contained variables that are known to be closely related. For example, lower anxiety was associated with higher self-efficacy and lower extraversion, as found in other studies (Qin & Li, 2024). Of course, it is likely that learners' aptitudes and L2 gains were influenced by other ID variables that we did not include in the present study. For example, we did not include the role of linguistic distance. While L1 distance has been shown to have a positive influence on L2 Dutch learning (Schepens et al., 2016), it would also be interesting to examine potential interaction effects between linguistic distance and affective and conative variables. While recognizing that it may be important to include additional variables, the five communities we identified could serve as a starting point for future studies taking a broader approach to aptitude.

Besides identifying communities, we also wanted to examine the relationships between communities. Our second RQ concerned which variables connected different communities together. Results highlighted the roles of L2 use, intended effort, and attitudes toward learning Dutch. The network loadings showed that these variables were connected to several other variables from different communities, indicating their relative importance within the network. From a network perspective, these variables may serve as causal bridges between communities (Fried, 2020). Our findings thus highlight the importance of motivated behavior (L2 use and intended effort) in connecting communities of variables related to motivation, proficiency level, and self-beliefs and emotion. It is also interesting to note that two bridge variables (L2 use and attitudes toward learning Dutch) were strongly related to proficiency level and LOR. This indicates that the role of these bridging variables might change if studying groups of learners at similar proficiency levels and LORs. For example, we would expect beginner learners to have low L2 use—because a certain level of proficiency is needed to use the L2—while for intermediate and advanced learners, L2 use is more likely to depend on their motivation, attitudes, and other characteristics. Additional analyses revealed further differential effects of L2 use in different contexts, whereby greater external pressure to learn Dutch (ought-to L2 self) was associated with more L2 use at work or school and at home, while greater internal motivation to learn Dutch (ideal L2 self) was associated with more L2 use with friends and free-time activities.

With our final RQ, we examined which ID variables were related to L2 gains. L2 gains were negatively associated with the ought-to L2 self, proficiency level, and LOR, and positively associated with the level of multilingualism. In meta-analyses of the L2 motivational self system, Al-Hoorie (2018) and Qi (2022) found no significant associations between the ought-to L2 self and measures of L2 achievement, and the ideal L2 self was a better predictor of outcomes. Our different findings could be explained by the study context: Participants were learners of Dutch as an L2, but research on language motivation is biased toward learners of English in foreign language contexts (Dörnyei & Al-Hoorie, 2017). For example, in meta-analyses, Al-Hoorie (2018) identified only 3 out of 32 studies that investigated a language other than English (LOTE), and Qi (2022) found moderating effects of the LOTE contexts on the ought-to L2 self. In the present study, the ideal L2 self appeared to have an indirect effect on L2 gains, via L2 use and proficiency level. The level of multilingualism had a positive influence on L2 gains. This has been found in several SLA studies (Sáfár & Kormos, 2008; Thompson, 2013) and generally supports SLA theories about the benefits of multilingualism for subsequent language learning (Cenoz, 2013; Hirosh & Degani, 2018). Similar to other studies (Botes et al., 2020; Dewaele et al., 2008), we also found that higher levels of multilingualism were associated with lower levels of anxiety. There were also interactions between multilingualism and the ought-to L2 self; in addition to both of these variables influencing L2 gains, learners with higher levels of multilingualism also had lower external pressures and obligations to learn Dutch.

Proficiency level had the strongest association with L2 gains. Participants with lower Dutch proficiency at the start of their course made more gains compared to those with higher proficiency. This association may be due to a variety of reasons. One reason could be that the C-test was more sensitive to measuring L2 gains at lower proficiency levels. While the C-test includes different aspects of L2 knowledge (e.g., vocabulary, grammar, sentence structure), it might be that we had fewer items to test aspects of L2 proficiency that are developed at higher proficiency levels. Only a small number of participants (16%) were enrolled in B2-level courses and higher, so we cannot be certain about the generalizability of our findings to higher proficiency levels. Another reason why participants with lower proficiency levels made greater gains could be that L2 development typically slows down: It takes more time, input, and instruction to get from A2 to B1 proficiency level than from 0 to A1, for example (Beningo et al., 2017). Nevertheless, the results of the EGA highlighted how much the proficiency level affected the relationships between ID variables and L2 learning outcomes. When looking at bivariate correlations (see Online Supporting Information D), there was a negative relationship between L2 gains and L2 use ($-.19$, $p > .001$). This relationship did not make sense without considering the effect of proficiency level; learners with lower proficiency levels had lower L2 use, but they made greater L2 gains compared to learners with higher proficiency levels. In a similar way, bivariate correlations showed a positive relationship between L2 gains and attitudes toward the Dutch community (.13, $p = .009$). The EGA showed that this relationship was moderated by how long participants had lived in the Netherlands, whereby participants with a longer LOR had more negative attitudes toward the Dutch community. In turn, the same variables that were dependent on proficiency level and LOR (L2 use and attitudes toward the Dutch community) were also related to ID variables across different communities, potentially acting as bridges between communities. All things considered, this points to a complex pattern of causal interactions between ID variables and L2 gains, determined largely by proficiency level.

The moderating effects of proficiency level and LOR suggest that it could be more meaningful to examine groups of learners with similar proficiency levels and LORs. Previous research has shown that different types of aptitudes are more strongly related to L2 gains at different proficiency levels (Artieda & Muñoz, 2016; Dunn & Iwaniec, 2022; Serafini, 2017). To account for this, some studies have demonstrated the benefits of taking a person-centered approach, using cluster analysis to group together learners based on ID variables (de Wilde & Lowie, 2024; Papi & Teimouri, 2014; Peng et al., 2022; Sparks et al., 2012). In our study, it is likely that the associations between L2 gains and ID variables were attenuated by the heterogeneity of our sample. The role of other variables might emerge more clearly if we focused on homogeneous groups of learners within our sample.

Inspired by Snow's (1992) theory of aptitude complexes, the main goal of our study was to explore the construct of aptitude from a holistic perspective, considering the joint influence of ID variables on L2 learning from across different domains. In the field of SLA, aptitude has long been defined and operationalized as a relatively stable set of cognitive abilities (see Li, 2016; Wen, 2021). Yet from both a theoretical and practical standpoint, there are no clear reasons why language aptitude should remain limited to cognitive abilities. The results of the present study showed that cognitive abilities were not strongly related to L2 learning outcomes or other ID variables. Rather, variables belonging to the communities of motivation and effort, and self-beliefs and emotion had a larger effect on L2 outcomes. Other multivariate studies of aptitude have also shown that L2 learning outcomes are dependent on a combination of cognitive, affective, and conative ID variables (Ehrman & Oxford, 1995; Gardner et al., 1997; Winke, 2013). Studying combinations of ID variables will likely better predict learning than studying a singular ID variable and/or domain (Ackerman, 2003). Our results also highlight the importance of the L2 context and the heterogeneity of the sample, as the relationships between aptitudes and L2 outcomes were partly dependent on learners' amount of L2 use, proficiency level, LOR, and attitudes toward the L2 community.

A broader view of aptitude inherently raises a number of questions, most notably regarding the number of ID variables to include and which ones. With open, complex systems, we cannot control or be aware of all elements that contribute to an individual learner's L2 development (Al-Hoorie et al., 2023). That said, we can draw on a wealth of empirical research on the role of IDs in SLA. Following Snow (1987) and Ackerman (2003), we could aim to identify whether the same patterns of relationships between aptitudes (A–A) and L2 learning outcomes (A–O) are found in different populations and contexts. The five communities of ID variables that emerged in our sample of L2 Dutch learners are largely consistent with previous SLA research. If the present study were replicated in a different learner population and context, we would expect very similar communities of ID variables to emerge. We would also expect to see similar moderating effects of proficiency level and LOR on the relationships between ID variables and L2 learning. If additional variables were included in analyses, we would expect them to join one of the five existing communities or potentially form new communities. For example, a motivational construct like goal orientation would likely join the motivation and effort community, while we might expect language analytic ability to join the proficiency level and use community. Regarding L2 learning gains, our results showed that learners' gains were directly related to ought-to L2 self and level of multilingualism and indirectly related to L2 use and attitudes toward the L2 community. We might expect these findings to be partially replicated in similar LOTE contexts. However, we may expect different findings for learners of L2 English, in foreign language contexts, and at different proficiency levels and stages of learning.

CONCLUSION

We conducted an exploratory study of aptitude with learners of L2 Dutch, using EGA to examine how combinations of aptitudes across different domains contribute to L2 learning. Our results highlight the benefits of taking a broader approach to language aptitude; we can only understand the effects of ID variables on L2 learning if we view them holistically, considering the interdependency of cognitive, affective, conative, and contextual processes. We argue that it is time for SLA scholars to reconnect with aptitude theory and question preexisting notions of aptitude. Instead of limiting our view of aptitude to cognitive abilities, we could expand our view and work toward a more holistic theory. This would essentially entail synthesizing SLA research findings and trying to establish patterns of commonalities between learners, while at the same time identifying how patterns may change in different contexts and populations. Building on Snow's (1992) theory of aptitude complexes, we conceptualize language aptitude as a complex system that emerges out of interactions between communities of closely related variables across cognitive, affective, and conative domains of L2 learning. We call on other SLA researchers to explore Snow's ideas and work toward an integrative, holistic theory of language aptitude.

AUTHOR CONTRIBUTIONS

Lani Freeborn: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); resources (lead); software (lead); visualization (lead); writing—original draft preparation (lead); writing—review and editing (lead). **Sible Andringa**: Conceptualization (supporting); project administration (supporting); supervision (lead); writing—original draft preparation (supporting); writing—review and editing (supporting). **Judith Rispens**: Conceptualization (supporting); project administration (supporting); supervision (supporting); writing—original draft preparation (supporting); writing—review and editing (supporting).

ACKNOWLEDGMENTS

We would like to thank Dirk Vet for programming the cognitive tests used in this study and Charlie Kirby for developing the website used for data collection. We would also like to thank the two anonymous reviewers for their comments and suggestions on how to improve the manuscript.

DATA AVAILABILITY STATEMENT

Data set is not publicly available yet. Materials and statistical analyses are available in Online Supporting Information.

ORCID

Lani Freeborn https://orcid.org/0000-0001-7818-3159

ENDNOTES

[1] See also https://osf.io/hjcvz/files/osfstorage

[2] We excluded L2 Dutch from the multilingualism score to avoid measurement overlap with the T1 Dutch C-test score.

[3] In our sample of 61 L1s, we were missing lexical distance values for 35 languages, morphological distance values for 24 languages, and both values for 20 languages.

[4] Dutch is most closely related to other West Germanic languages including Frisian, German, Luxembourgish, and English. In the Netherlands, it is commonly acknowledged that learners with prior knowledge of Germanic languages experience positive transfer, and many language schools even offer "fast-track" Dutch language courses for L1 German speakers.

REFERENCES

Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist*, *38*, 85–93. https://doi.org/10.1207/S15326985EP3802_3

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245. https://doi.org/10.1037/0033-2909.121.2.219

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*, 163–181. https://doi.org/10.1037/a0015719

Al-Hoorie, A. H. (2018). The L2 motivational self system: A meta-analysis. *Studies in Second Language Learning and Teaching*, *8*, 721–754. https://doi.org/10.14746/ssllt.2018.8.4.2

Al-Hoorie, A. H., Hiver, P., Larsen-Freeman, D., & Lowie, W. (2023). From replication to substantiation: A complexity theory perspective. *Language Teaching*, *56*, 276–291. https://doi.org/10.1017/S0261444821000409

Artieda, G., & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences*, *50*, 42–48. https://doi.org/10.1016/j.lindif.2016.06.023

Beningo, V., de Jong, J., & van Moere, A. (2017). *How long does it take to learn a language? Insights from research on language learning*. *Global Scale of English Research Series*. Pearson. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/SupportingDocs/GSE-learning-hours.pdf

Biedroń, A. (2011). Personality factors as predictors of foreign language aptitude. *Studies in Second Language Learning and Teaching*, *1*, 467–489. https://doi.org/10.14746/ssllt.2011.1.4.2

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*, 354–369. https://doi.org/10.1177/1073191112446655

Botes, E., Dewaele, J., & Greiff, S. (2020). The power to improve: Effects of multilingualism and perceived proficiency on enjoyment and anxiety in foreign language learning. *European Journal of Applied Linguistics*, *8*, 279–306. https://doi.org/10.1515/eujal-2020-0003

Botes, E., van der Westhuizen, L., Dewaele, J.-M., MacIntyre, P., & Greiff, S. (2022). Validating the short-form Foreign Language Classroom Anxiety Scale. *Applied Linguistics*, *43*, 1006–1033. https://doi.org/10.1093/applin/amac018

Carroll, J. B. (1964). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87–136). University of Pittsburgh Press.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Newbury House.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*. Psychological Corporation.

Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*, *46*, 71–86. https://doi.org/10.1017/S0261444811000218

Christensen, A. P., Garrido, L. E., & Golino, H. (2023). Unique variable analysis: A network psychometrics method to detect local dependence. *Multivariate Behavioral Research*, *58*, 1165–1182. https://doi.org/10.1080/00273171.2023.2194606

Christensen, A. P., & Golino, H. (2021a). Estimating the stability of psychological dimensions via bootstrap Exploratory Graph Analysis: A Monte Carlo simulation and tutorial. *Psych*, *3*, 479–500. https://doi.org/10.3390/psych3030032

Christensen, A. P., & Golino, H. (2021b). On the equivalency of factor and network loadings. *Behavior Research Methods*, *53*, 1563–1580. https://doi.org/10.3758/s13428-020-01500-6

Cochran, J. L., McCallum, R. S., & Bell, S. M. (2010). Three A's: How do attributions, attitudes, and aptitude contribute to foreign language learning? *Foreign Language Annals*, *43*, 566–582. https://doi.org/10.1111/j.1944-9720.2010.01102.x

Dewaele, J.-M., Petrides, K. V., & Furnham, A. (2008). Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning*, *58*, 911–960. https://doi.org/10.1111/j.1467-9922.2008.00482.x

De Wilde, V., & Lowie, W. (2024). The forest and the trees: Investigating groups and individuals in longitudinal second language English speaking development. *Language Learning*, Advance online publication. https://doi.org/10.1111/lang.12698

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum. https://doi.org/10.4324/9781410613349

Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic systems perspective. In E. Macaro (Ed.), *Continuum companion to second language acquisition* (pp. 247–267). Continuum.

Dörnyei, Z., & Al-Hoorie, A. (2017). The motivational foundation of learning languages other than global English: Theoretical issues and research directions. *Modern Language Journal*, *101*, 455–468. https://doi.org/10.1111/modl.12408

Dunn, K., & Iwaniec, J. (2022). Exploring the relationship between second language learning motivation and proficiency: A latent profiling approach. *Studies in Second Language Acquisition*, *44*, 967–997. https://doi.org/10.1017/S0272263121000759

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*, 290–325. https://doi.org/10.1191/0265532206lt330oa

Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal*, *79*, 67–89. https://doi.org/10.2307/329394

Epskamp, S., & Fried, E. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*, 617–634. https://doi.org/10.1037/met0000167

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149. https://doi.org/10.3758/BF03203267

Fried, E. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*, 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Gardner, R. C. (2004). Attitude/Motivation Test Battery: International AMTB research project. https://languageattrition.org/wp-content/uploads/2019/06/robert-gardner-attitude-and-motivational-test-battery.pdf

Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a full model of second language learning: An empirical investigation. *Modern Language Journal*, *81*, 344–362. https://doi.org/10.2307/329310

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Golino, H., Moulder, R., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Nesselroade, J., Sadana, R., Thiyagarajan, J. A., & Boker, S. M. (2020a). Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables. *Multivariate Behavioral Research*, *56*, 874–902. https://doi.org/10.1080/00273171.2020.1779642

Golino, H. F., & Christensen, A. P. (2024). EGAnet: Exploratory graph analysis—A framework for estimating the number of dimensions in multivariate data using network psychometrics (R package version 2.3.0). https://doi.org/10.32614/CRAN.package.EGAnet

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE*, *12*, e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H. F., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020b). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*, 292–320. https://doi.org/10.1037/met0000255

Graham, S. (2021). Aptitude or motivation: Which is the better predictor of successful language learning? In E. Macaro, & R. Woore (Eds.), *Debates in second language education* (pp. 106–121). Routledge. https://doi.org/10.4324/9781003008361

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*, 665–703. https://doi.org/10.1111/lang.12018

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory-based approach to the measurement of foreign language learning ability: The Canal-F Theory and Test. *Modern Language Journal*, *84*, 390–405. https://doi.org/10.1111/0026-7902.00076

Hirosh, Z., & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin & Review*, *25*, 892–916. https://doi.org/10.3758/s13423-017-1315-7

Hiver, P., & Al-Hoorie, A. H. (2020). Reexamining the role of vision in second language motivation: A preregistered conceptual replication of You, Dörnyei, and Csizér (2016). *Language Learning*, *70*, 48–102. https://doi.org/10.1111/lang.12371

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*, 125–132. https://doi.org/10.2307/327317

Kliesch, M., & Pfenninger, S. E. (2021). Cognitive and socioaffective predictors of L2 microdevelopment in late adulthood: A longitudinal intervention study. *Modern Language Journal*, *105*, 237–266. https://doi.org/10.1111/modl.12696

Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 131–152). John Benjamins.

Language Learning and Testing Foundation. (2025). https://lltf.net/

Lee, S. (2020). Examining the roles of aptitude, motivation, strategy use, language processing experience, and gender in the development of the breadth and depth of EFL learners' vocabulary knowledge. *SAGE Open*, *10*(4). https://doi.org/10.1177/2158244020977883

Lee-Ellis, S. (2009). The development and validation of a Korean C-test using Rasch analysis. *Language Testing*, *26*, 245–274. https://doi.org/10.1177/0265532208101007

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, *144*, 394–425. https://doi.org/10.1037/bul0000142

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*, 801–842. https://doi.org/10.1017/S027226311500042X

Li, S. (2024). Individual differences and task-based language teaching: Theory, research, and practice. In S. Li (Ed.), *Individual differences and task-based language teaching* (pp. 10–49). John Benjamins.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–566. https://doi.org/10.1111/lang.12011

MacIntyre, P. D., Babin, P. A., & Clément, R. (1999). Willingness to communicate: Antecedents and consequences. *Communication Quarterly*, *47*, 215–229. https://doi.org/10.1080/01463379909370135

Maris, G., Bechger, T., Koops, J., & Partchev, I. (2024). *dexter: Data management and analysis of tests* (R package version 1.5.1). https://dexter-psychometrics.github.io/dexter/

McKay, T. H., Teimouri, Y., Sağdıç, A., Salen, B., Reagan, D., & Malone, M. E. (2021). The cagey C-test construct: Some evidence from a meta-analysis of correlation coefficients. *System*, *99*, 102526. https://doi.org/10.1016/j.system.2021.102526

Meara, P. (2005). *Llama language aptitude tests: The manual*. https://llamatests.org/docs/llama_manual.pdf

Mozgalina, A., & Ryshina-Pankova, M. (2015). Meeting the challenges of curriculum construction and change: Revision and validity evaluation of a placement test. *Modern Language Journal*, *99*, 346–370. https://doi.org/10.1111/modl.12217

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*, 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. *Language Learning*, *64*, 493–525. https://doi.org/10.1111/lang.12065

Paxinou, E., Kalles, D., Panagiotakopoulos, C. T., Sgourou, A., & Verykios, V. S. (2021). An IRT-based approach to assess the learning gain of a virtual reality lab students' experience. *Intelligent Decision Technologies*, *15*, 487–496. https://doi.org/10.3233/IDT-200216

Peng, H., Jager, S., & Lowie, W. (2022). A person-centred approach to L2 learners' informal mobile language learning. *Computer Assisted Language Learning*, *35*, 2148–2169. https://doi.org/10.1080/09588221.2020.1868532

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Qi, D. (2022). Examining the L2 motivational self system through meta-analysis (Publication No. 29160892) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations & Theses.

Qin, F., & Li, W. (2024). Personality traits and foreign language anxiety: The mediating role of self-efficacy. *Journal of Psychology in Africa*, *34*, 360–366. https://doi.org/10.1080/14330237.2024.2371709

Qualtrics. [Computer software]. (2022). https://www.qualtrics.com

R Core Team. (2023). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raatz, U., & Klein-Braley, C. (2002). Introduction to language testing and to C-tests. In J. A. Coleman, R. Grotjahn, & U. Raatz (Eds.), *University language testing and the C-test* (pp. 75–87). Bochum.

Raven, J., & Raven, J. (2003). Raven progressive matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 223–237). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0153-4_11

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*, 368–392. https://doi.org/10.1177/026765830101700405

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. https://doi.org/10.1017/S0267190505000036

Rogers, V., Meara, P., & Rogers, B. (2023). Testing language aptitude: LLAMA evolution and refinement. In Z. Wen, P. Skehan, & R. L. Sparks (Eds.), *Language aptitude theory and practice* (pp. 47–72). Cambridge University Press.

Ryan, S. (2009). Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. In Z. Dornyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 120–143). Multilingual Matters. https://doi.org/10.21832/9781847691293-007

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *International Review of Applied Linguistics in Language Teaching*, *46*, 113–136. https://doi.org/10.1515/IRAL.2008.005

Sasaki, M. (2012). The Modern Language Aptitude Test (paper-and-pencil version). *Language Testing*, *29*, 315–321. https://doi.org/10.1177/0265532211434015

Schepens, J. J., van der Slijk, F., & van Hout, R. (2016). L1 and L2 distance effects in learning L3 Dutch. *Language Learning*, *66*, 224–256. https://doi.org/10.1111/lang.12150

Schnoor, B., Hartig, J., Klinger, T., Naumann, A., & Usanova, I. (2023). Measuring the development of general language skills in English as a foreign language: Longitudinal invariance of the C-test. *Language Testing*, *40*, 796–819. https://doi.org/10.1177/02655322231159829

Serafini, E. (2017). Exploring the dynamic long-term interaction between cognitive and psychosocial resources in adult second language development at varying proficiency. *Modern Language Journal*, *101*, 369–390. https://doi.org/10.1111/modl.12400

Serafini, E., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*, 607–646. https://doi.org/10.1017/S0272263115000327

Singleton, D. (2017). Language aptitude: Desirable trait or acquirable attribute? *Studies in Second Language Learning and Teaching*, *7*, 89–103. https://doi.org/10.14746/ssllt.2017.7.1.5

Snow, R. E. (1963). *Effects of learner characteristics in learning from instructional films* (Publication No. 6404928) [Doctoral dissertation]. ProQuest.

Snow, R. E. (1987). Aptitude complexes. In R. E. Snow, & M. J. Farr (Eds.) *Aptitude, learning, and instruction* (pp. 11–34). Routledge.

Snow, R. E. (1989a). Aptitude, instruction, and individual development. *International Journal of Educational Research*, *13*, 869–881. https://doi.org/10.1016/0883-0355(89)90070-0

Snow, R. E. (1989b). Cognitive–conative aptitude interactions in learning. In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.) *Abilities, motivation and methodology: The Minnesota Symposium on Learning and Individual Differences* (pp. 435–474). Lawrence Erlbaum.

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, *27*, 5–32. https://doi.org/10.1207/s15326985ep2701_3

Snow, R. E., & Jackson, D. N. (1997). Individual differences in conation: Selected constructs and measures. In H. F. O'Neil, Jr., & M. Drillings (Eds.), *Motivation: Theory and research* (pp. 71–99). Routledge.

Sparks, R. L., Patton, J., & Ganschow, L. (2012). Profiles of more and less successful L2 learners: A cluster analysis study. *Learning and Individual Differences*, *22*, 463–472. https://doi.org/10.1016/j.lindif.2012.03.009

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2011). Subcomponents of second-language aptitude and second-language proficiency. *Modern Language Journal*, *95*, 253–273. https://doi.org/10.1111/j.1540-4781.2011.01176.x

Stansfield, C. W., & Reed, D. J. (2004). The story behind the Modern Language Aptitude Test: An interview with John B. Carroll (1916–2003). *Language Assessment Quarterly*, *1*, 43–56. https://doi.org/10.1207/s15434311laq0101_4

Thompson, A. S. (2013). The interface of language aptitude and multilingualism: Reconsidering the bilingual/multilingual dichotomy. *Modern Language Journal*, *97*, 685–701. https://doi.org/10.1111/j.1540-4781.2013.12034.x

Thompson, A. S., & Lee, J. (2012). Anxiety and EFL: Does multilingualism matter? *International Journal of Bilingual Education and Bilingualism*, *16*, 730–749. https://doi.org/10.1080/13670050.2012.713322

Vet, D. J. (n.d.). *Experiment designer* [Computer software]. https://www.fon.hum.uva.nl/dirk/ed.php

Wechsler, D. (1997). *WAIS-3: Wechsler adult intelligence scale: Technical manual*. Psychological Corporation.

Wen, Z. E. (2021). Language aptitudes. In T. Gregersen, & S. Mercer (Eds.), *The Routledge handbook of the psychology of language learning and teaching* (pp. 391–405). Routledge. https://doi.org/10.4324/9780429321498

Wen, Z. E. (2022). Working memory. In H. Mohebbi, & C. Coombe (Eds.), *Research questions in language education and applied linguistics* (pp. 279–284). Springer. https://doi.org/10.1007/978-3-030-79143-8_50

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *Modern Language Journal*, *97*, 109–130. https://doi.org/10.1111/j.1540-4781.2013.01428.x

Yilmaz, Y., & Granena, G. (2019). Cognitive individual differences as predictors of improvement and awareness under implicit and explicit feedback conditions. *Modern Language Journal*, *103*, 686–702. https://doi.org/10.1111/modl.12587

You, C., Dörnyei, Z., & Csizér, K. (2016). Motivation, vision, and gender: A survey of learners of English in China. *Language Learning*, *66*, 94–123. https://doi.org/10.1111/lang.12140

SUPPORTING INFORMATION

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

How to cite this article: Freeborn, L., Andringa, S., & Rispens, J. (2025). Aptitude complexes: Expanding our view of language aptitude. *Modern Language Journal*, *109*, 775–799. https://doi.org/10.1111/modl.70002