Appendix S1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Study | Implicit learning aptitude task | Language type | Outcome construct | Outcome measure | Finding | *r* (ILA, outcome) |
| Antoniou, Ettlinger, & Wong (2016) | TOL  (accuracy-based *z* score reflecting improvement on the second repetition of sequences, p. 7) | Artificial | Grammar knowledge  (rule based/simple and associative/complex) | A modified wug test (two-alternative forced choice recognition test) | **Simple structure:**  a significant correlation was found between procedural memory and grammar learning;  **Complex structure:**  no such significant relationship was found (correlation with declarative memory instead) | Rule-based/simple structure:  .531 (*p* < .001)  Associative/complex structure:  .14 (*p* > .05, ns) |
| Brooks & Kempe (2013) | Verbal auditory statistical learning test  (measuring the ability to detect adjacent dependencies; test adapted from Misyak & Christiansen, 2007) | Natural | Vocabulary and grammar knowledge | Production task (grammar knowledge); picture naming task (vocabulary knowledge) | The effect of auditory sequence learning on case-marking performance was mediated by metalinguistic awareness: auditory sequence learning predicted metalinguistic awareness, which in turn predicted the development of case marking | Without metalinguistic knowledge as a predictor:  ~learning of case marking: β = .24 (*p* < .05);    With metalinguistic knowledge as a predictor:  ~ learning of case marking: β = .12 (*p* > .05, ns) |
| Buffington & Morgan-Short (2018) | TOL (*z* score from the average total time in the second administration of the task); WPT; ASRT | Artificial | Grammar knowledge (rule based/simple and associative/complex) | A modified wug test  (two-alternative forced choice recognition test) | **Reliability:**  WPT and ASRT showed acceptable reliability. TOL reliability was low.  **(Lack of) convergent validity:**  The three procedural memory measures did not correlate with each other or correlated negatively *r*(WPT, ASRT) = -.37, *p* <.05) | Rule-based/simple structure:  ~WPT: .47 (*p* <.05)  ~ASRT: .04  ~TOL: -.12  Associative/complex structure:  ~WPT: -.18 (*p* >.10)  ~ASRT: .04  ~TOL: -.12 |
| Ettlinger, Bradlow, & Wong (2014) | TOL (accuracy-based *z* score reflecting improvement on the second repetition of sequences, pp. 813-814) | Artificial | Grammar knowledge (rule based/simple and associative/complex) | A wug test | A significant correlation between procedural memory and performance on simple words was found. However, no such procedural memory-performance association exists with complex words. | Rule-based/simple structure:  ρ = .52 (p < .001)  Associative/complex structure:  ρ = .47 (*p* = .10, ns) |
| Faretta-Stutenberg & Morgan-Short (2018) | ASRT (difference in reaction time for correctly responded pattern versus random trials); the dual-task version of WPT | natural | Grammar knowledge | Untimed written grammaticality judgment test | At home students: Procedural memory did not significantly predict performance gains.    Study abroad students: Procedural learning ability was a statistically significant, unique predictor of performance gains. | At home:  *β =* .365 (*p* > .05, ns)    Study abroad:  *β =* .855 (*p* < .01) |
| Granena (2013) | SRT;  LLAMA D | natural | Grammar knowledge | Metalinguistic knowledge test; word monitoring task | Sequence learning ability moderated scores on structures involving grammatical agreement relations in both early and late second language learners | Early group, agreement structures, MLK ~ LLAMA D: .41 (*p* < .01);    Late group, agreement structures, WMT ~SRT: .36 (*p =* .01) |
| Granena (2019) | LLAMA D; ALTM Synonym (a rate score combining response time and accuracy);  SRT (a facilitation score was derived as the difference in median reaction times in the last sequential block and the last random block) | Natural | Speaking (CAF measures were used) | Oral production task | ALTM Synonym and LLAMA D loaded on the same component, labeled as implicit memory ability. SRT loaded onto a different component, labeled as implicit learning aptitude. Implicit memory predicted speech fluency and, for learners with high implicit learning ability, it also predicted lexical complexity. | Fluency:  *R2* = .08  (model with implicit memory ability)    Lexical complexity:  *R2* = .11  (model with implicit memory ability × implicit learning aptitude interaction) |
| Hamrick (2015) | SRT | Semi-artificial | Grammar knowledge | Recognition task | Procedural memory ability predicted performance on the delayed task | Immediate posttest:  -.20,(*p* = .29, ns)  Delayed posttest:  .43 (*p* < .01) |
| Linck et al. (2013) | SRT | Natural language | Listening;  Reading | ILR scale | High-level attainment was related to implicit learning aptitude | Reading: .51 (*p* < .05)  Listening: .40 (*p* < .05) |
| McDonough & Trofimovich (2016) | Verbal auditory statistical learning test (measuring the ability to detect non-adjacent dependencies; test adapted from Gomez, 2002) | Artificial | Pattern learning (grammar knowledge) | Picture identification task | Better performance on pattern learning was associated with greater statistical learning ability | Novel language with exposure:  .51 (*p* < .05)  Familiar language with exposure:  .35 (*p* < .05)  Familiar language without exposure:  .05 (*p* > .05) |
| Misyak & Christiansen (2012) | Verbal auditory adjacent and non-adjacent statistical learning tasks (non-adjacent task modeled after Gomez, 2002; adjacent task adapted from Friederici, Steinhauer, & Pfeifer, 2002) | Natural | Grammar knowledge | Self-paced reading task | For the comprehension test about ambiguities involving phonological typicality, only adjacent-statistical learning showed a strong positive relationship to language comprehension; for the comprehension test about subject-object relative clauses, only nonadjacent-statistical learning was strongly related to language comprehension | Adjacent statistical learning  ~ language comprehension (ambiguities involving phonological typicality): β = .42, p < .05  Nonadjacent statistical learning  ~ language comprehension (subject-object relative clauses): β = .38, p < .05 |
| Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong (2014) | TOL (ITT)average proportional change scores were computed from the first to the final trial for each set of trials (RT on the first trial - RT on the final trial)/RT on the first trial;  WPT | Artificial | Syntactic knowledge | Timed aural grammaticality judgment test | Positive relationship between procedural learning ability and development in **later** stages of acquisition. | Early stages:  TOL: .187 ns  WPT: .281, ns    Later stages:  TOL: .588 (*p* < .01)  WPT: .557 (*p* < .01) |
| Pili-Moss et al. (2019) | TOL (ITT) average proportional change scores were computed from the first to the final trial for each set of trials (RT on the first trial - RT on the final trial)/RT on the first trial; WPT | Artificial | Sentence comprehension and production | Sentence comprehension (accuracy and RT-based) & production (accuracy-based) | A positive, medium, statistically significant relationship was found between declarative learning ability and comprehension accuracy.    Procedural learning ability had a positive, medium, significant effect on automatization, whereas declarative learning ability had a positive, small effect that was not statistically significant. | Comprehension accuracy and declarative learning ability: *R2* = .30 (*p* < .001)      Automatization and procedural learning ability: *R2* = .30 (*p* < .01) |
| Suzuki (2017) | TOL (RT measures; ITT, MET, OST*)*, the average proportional change scores were computed from the first to the final trial for each set of trials (RT on the first trial - RT on the final trial)/RT on the first trial.) | Semi-artificial | Grammar & vocab | RT-based oral production tests in DMDX (three tests: vocab test; rule-application test; present progressive test) | Procedural learning ability measured by the Tower of London task was significantly associated with RT, but not with CV, in the short-spacing learning condition group only | -.38 < *r < -*.51 (*p* < .05), with MET (a RT measure of TOL) |
| Suzuki & DeKeyser (2015) | SRT | Natural | Grammar knowledge | Elicited imitation with built-in word monitoring task; metalinguistic knowledge test | EIM task performance, measured by GSI, was correlated positively with SRT scores only among L2 speakers with long LOR | .43 (*p* = .065) |
| Suzuki & DeKeyser (2017) | SRT | Natural | Grammar knowledge | Visual world eye tracking; word monitoring task; self-paced reading | No significant relationship was found between IK aptitude and performance on language knowledge measures | ns |
| Tagarelli, Ruiz, Vega, & Rebuschat (2016) | SRT & ASRT | semi-artificial | Grammar knowledge | Untimed auditory grammaticality judgment test | **Incidental group:** strong negative correlation between the SRTz (*z* scores from SRT and ASRT task); no correlation between RSpan and *d'* scores.  **Instructed group:** no correlation between the cognitive variables and overall GJT performance | -.586 (*p* = .003) |
| Yi (2018) | SRT, LLAMA D (later excluded) | Natural | Statistical sensitivity to collocation frequency and contingency | Phrasal acceptability judgement task | Implicit language aptitude (SRT) was predictive of L1 speakers’ reaction time but not for L2 speakers;  Implicit language aptitude (SRT) was found not associated with judgment accuracy | L1\*implicit language aptitude: *b* = 1 - exp(-.06), *p* < .01 |

*Notes:* ASRT = alternating serial reaction time task; ITT = initial thinking time; MET = movement execution time; OST = overall solution time; TOL = Tower of London; WPT = weather prediction task

Appendix S2

**Metalinguistic Knowledge Test**

**Scoring Guide & Answer Key**

**General principles**

1. To score a point for explanation, participants need to explain why they make such corrections; providing the rule and metalinguistic terms is often evidence of metalinguistic knowledge.
2. Mere description of the correction does not suffice an explanation.
3. If one part of an explanation violates the rule/contradicts another part of the explanation, it should be scored as incorrect.
4. Do not penalize for non-relevant mistakes/errors. For example, if the verb complement error “needs asking” is corrected to “need to ask”, the learner should still get the point, as the original error was corrected, despite the learner introducing a 3rd-person singular error. Learners vary in how much of the sentence they reproduce during corrections, which is not strictly relevant to what the item is targeting.

**1. Third-person – s**

Full explanation should consist of = because the noun/subject/“the name

of the word appearing in the sentence” is singular, verb + s should be added to the verb/not plural.

*Participants will have to mention:*

(1) the (noun / subject) is (singular / third-person / third person singular); AND

(2) the -s form of the verb needs to be used / -s needs to be added to the verb

**2. Mass/count nouns**

Full explanation should consist of = mention of specific terminology, such as countable/uncountable or can be plural/cannot be plural.

*Participants will have to mention:*

(1) the noun (is not countable/ is uncountable /cannot be plural / cannot have -s added to it)

**3. Be passive**

Full explanation should consist of = mention of the subject is a receiver of an action or not an active subject, and thus requires a passive verb form.

*Participants will have to mention*:

(1) the subject is (the receiver of an action / not an active subject); OR

(2) passive voice is needed here.

**5. Embedded questions**

Full explanation should consist of = the position of the subject and an auxiliary verb or the verb “to be”/ a verb have to follow affirmative sentence order, not the word order as in direct questions.

*Participants will have to mention:*

(1) follow word order of a statement; OR

(2) follow SVO [word order]; OR

(3) inversion is not needed.

**6. Comparatives**

Full/required explanation should consist of = add ***–er*** to one-syllable adjectives to make a comparison or add ***more*** in front of adjectives with more than two syllables.

*Participants will have to mention:*

(1) ***-er*** is used for (short words / words with one syllables); OR

(2) ***more*** is used for (long words / words with two or more syllables); OR

(3) double marking is not needed; OR

**7. Gerund/Infinitives**

Full/required explanation should consist of = the infinitive form is needed after the initial verb in the sentence. Only one verb takes inflection in English clauses.

*Participants will have to mention*:

(1) [after to / after certain verbs] (infinitive/base/dictionary form) of the verb is needed; OR

(2) [after to / after certain verbs] (-ing form/gerund form) is not needed

**Notes:**

(alternatives / other options acceptable)

[optional part of the response]

|  |  |  |
| --- | --- | --- |
| Item | Stem | Key |
| 1 | William lives in Ann Arbor but **work** in East Lansing.  (third-person *-s*) | a) work  b) works (1 pt)  c) *Participants have to mention:*  (1) the (noun / subject) is (singular / third-person / third person singular); AND  (2) the -s form of the verb needs to be used / -s needs to be added to the verb  (1 pt) |
| 2 | Martin’s presentation **was post** on Facebook by his classmates.  (passive) | a) post  b) posted (1 pt)  c) *Participants have to mention*:  (1) the subject is (the receiver of an action / not an active subject); OR  (2) passive voice is needed here. (1 pt) |
| 3 | I asked Alan when **is he** going to play basketball.  (embedded question) | a) is he  b) he is (1 pt) [he was]  c) *Participants have to mention:*  (1) follow word order of a statement; OR  (2) follow SVO [word order]; OR  (3) inversion is not needed. (1 pt) |
| 4 | Diane wants to buy new **furnitures** and find the cat another home.  (mass/count noun) | a) furnitures  b) furniture (1 pt)  c) *Participants have to mention:*  (1) the noun (is not countable/ is uncountable /cannot be plural / cannot have -s added to it) (1 pt) |
| 5 | Lucy feels she **needs asking** for help with learning English.  (to/-ing verb complement) | a) asking  b) to ask (1 pt)  c) *Participants have to mention*:  (1) [after to / after certain verbs] (infinitive/base/dictionary form) of the verb is needed; OR  (2) [after to / after certain verbs] (-ing form/gerund form) is not needed (1pt) |
| 6 | People think he is **more nicer** and more intelligent than Peter.  (comparative adj.) | a) more nicer  b) nicer (1 pt)  c) *Participants will have to mention:*  (1) ***-er*** is used for (short words / words with one syllables); OR  (2) ***more*** is used for (long words / words with two or more syllables); OR  (3) double marking is not needed (1 pt) |
| 7 | We **were question** by the immigration officer at the airport.  (passive) | a) were question  b) were questioned (1 pt)  c) *Participants have to mention*:  (1) the subject is (the receiver of an action / not an active subject); OR  (2) passive voice is needed here. (1 pt) |
| 8 | Everybody **know** that teenagers like to play computer games.  (third-person *-s*) | a) know  b) knows (1 pt)  c) *Participants have to mention:*  (1) the (noun / subject) is (singular / third-person / third person singular); AND  (2) the -s form of the verb needs to be used / -s needs to be added to the verb  (1 pt) |
| 9 | She wondered why **did her boyfriend come** late for dinner.  (embedded question) | a) did her boyfriend come  b) why her boyfriend came (1 pt)  c) *Participants have to mention:*  (1) follow word order of a statement; OR  (2) follow SVO [word order]; OR  (3) inversion is not needed. (1 pt) |
| 10 | The temperature is **more higher** now in winter than it was ten years ago.  (comparative adj.) | a) more higher  b) higher (1 pt)  c) *Participants will have to mention:*  (1) ***-er*** is used for (short words / words with one syllables); OR  (2) ***more*** is used for (long words / words with two or more syllables); OR  (3) double marking is not needed (1 pt) |
| 11 | On the weekend she **asks playing** video games for an hour.  (to/-ing verb complement) | a) asks playing  b) asks to play (1 pt)  c) *Participants will have to mention*:  (1) [after to / after certain verbs] (infinitive/base/dictionary form) of the verb is needed; OR  (2) [after to / after certain verbs] (-ing form/gerund form) is not needed (1 pt) |
| 12 | Adam will get help on the **homeworks**, so he is not worried.  (mass/count noun) | a) homeworks  b) homework (1 pt)  c) *Participants will have to mention:*  (1) the noun (is not countable/ is uncountable /cannot be plural / cannot have -s added to it) (1pt) |

Appendix S3

Two-Factor CFA Model Parameter Estimates

|  |  |  |
| --- | --- | --- |
| Path | Standardized Est. | P value |
| Explicit → |  |  |
| UWGJT | 1.000 | - |
| UAGJT | 0.163 | 0.122 |
| MKT | 0.526 | 0.000 |
| Implicit → |  |  |
| EI | 0.587 | - |
| SPR | 0.225 | 0.034 |
| WMT | 0.054 | 0.611 |
| TWGJT | 0.833 | 0.000 |
| TAGJT | 0.528 | 0.000 |
| OP | 0.516 | 0.000 |

Three-Factor CFA Model Parameter Estimates

|  |  |  |
| --- | --- | --- |
| Path | Standardized Est. | P value |
| Explicit → |  |  |
| UWGJT | 1.000 | - |
| UAGJT | 0.162 | 0.124 |
| MKT | 0.527 | 0.000 |
| Automatized  Explicit → |  |  |
| EI | 0.586 | - |
| OP | 0.517 | 0.000 |
| TWGJT | 0.831 | 0.000 |
| TAGJT | 0.528 | 0.000 |
| Implicit → |  |  |
| SRT | 1.000 | - |
| WMT | 0.183 | 0.092 |