**Supplementary Material**

1. **Effect sizes contributed by each primary study based on oral proficiency traits**

|  |  |  |  |
| --- | --- | --- | --- |
| **Study name** | **Outcome** | **Hedges’s g** | **S. E.** |
| Abrams (2003) | Syntactic complexity | 0.02 | 0.18 |
|  | Fluency | -0.24 | 0.61 |
|  | Lexical richness | -0.27 | 0.18 |
|  | Lexical density | -0.29 | 0.18 |
| AbuSeileek (2007) | Overall | 1.35 | 0.19 |
| Ahn (2006) | Overall | 0.12 | 0.34 |
| Alastuey (2010) | Pronunciation | 0.52 | 0.22 |
| Alastuey (2011) | Overall | 0.30 | 0.40 |
| Blake (2009) | Overall | 0.25 | 0.25 |
| Blake *et al*. (2008) | Overall | -0.06 | 0.21 |
| Chang (2007) | Overall | 0.07 | 0.32 |
|  | Accuracy | -0.98 | 0.30 |
| Chang (2008) | Overall | 0.29 | 0.20 |
| Chen (2008) | Overall | 1.49 | 0.44 |
| Huang & Hung (2010) | Fluency | 0.66 | 0.37 |
|  | Lexical richness | 0.81 | 0.37 |
|  | Syntactic complexity | 0.38 | 0.36 |
| Kost (2004) | Overall | -0.03 | 0.18 |
| Li (2008) | Overall | 0.73 | 0.24 |
| Lord (2008) | Pronunciation | 0.94 | 0.36 |
| Payne & Whitney (2002) | Overall | 0.33 | 0.20 |
| Pyun (2003) | Syntactic complexity | -0.90 | 0.44 |
|  | Accuracy | 0.81 | 0.34 |
| Sanders (2005) | Overall | -0.29 | 0.26 |
| Satar & Özdener (2008) | Overall | 1.61 | 0.21 |
| Sequeira (2009) | Overall | 0.85 | 0.28 |
| Sun (2012) | Accuracy | -0.30 | 0.13 |
|  | Syntactic complexity | -0.27 | 0.12 |
|  | Fluency | -0.35 | 0.13 |
|  | Pronunciation | -0.21 | 0.12 |
| Volle (2005) | Pronunciation | 0.27 | 0.18 |
|  | Accuracy | -0.01 | 0.32 |
|  | Overall | 0.61 | 0.33 |
| Wang (2010) | Overall | 0.46 | 0.28 |
|  | Fluency | -1.88 | 0.23 |
|  | Lexical complexity | 1.71 | 0.32 |
|  | Syntactic complexity | 2.51 | 0.37 |
|  | Accuracy | -1.63 | 0.32 |
| Xiao (2007) | Accuracy | 2.07 | 0.54 |
|  | Syntactic complexity | 1.63 | 0.50 |
|  | Fluency | 1.78 | 0.51 |
| Yang (2006) | Lexical density | -0.01 | 0.24 |
|  | Lexical richness | 0.13 | 0.24 |
|  | Fluency | 0.01 | 0.24 |
|  | Syntactic complexity | 0.35 | 0.25 |
| Zheng (2010) | Syntactic complexity | -0.78 | 0.23 |
|  | Fluency | -1.32 | 1.72 |
|  | Lexical complexity | 0.61 | 0.33 |
|  | Accuracy | -0.97 | 0.76 |

1. **Research design and type of data for each primary study**

|  |  |  |
| --- | --- | --- |
| **Study** | **Research design** | **Data type** |
| Abrams (2003) | Pre-postest control vs. experiment | Naturalistic |
| AbuSeileek (2007) | Posttest only control (F2F) vs. experiment | Elicited |
| Ahn (2006) | One group only pre-posttest | Elicited |
| Alastuey (2010) | One group only pre-posttest | Elicited |
| Alastuey (2011) | Pre-postest control (F2F) vs. experiment | Elicited |
| Blake (2009) | Pre-postest control (F2F) vs. experiment | Elicited |
| Blake *et al*. (2008) | Posttest only control (F2F) and experiment | Elicited |
| Chang (2008) | Pre-postest control (F2F) vs. experiment | Elicited |
| Chang (2007) | Posttest only control (F2F) vs. experiment | Naturalistic |
| Chen (2008) | Pre-postest control vs. experiment | Elicited |
| Huang & Hung (2010) | Pre-postest control vs. experiment | Elicited |
| Kost (2004) | Pre-postest control (F2F) vs. experiment | Elicited |
| Li (2008) | One group only pre-posttest | Elicited |
| Lord (2008) | One group only pre-posttest | Elicited |
| Payne & Whitney (2002) | Pre-postest control (F2F) vs. experiment | Elicited |
| Pyun (2003) | Posttest only control (F2F) and experiment | Naturalistic |
| Sanders (2005) | Posttest only control (F2F) vs. experiment | Elicited |
| Satar & Özdener (2008) | Pre-postest control vs. experiment | Elicited |
| Sequeira (2009) | Pre-postest control (F2F) vs. experiment | Elicited |
| Sun 2012 | One group only pre-posttest | Naturalistic |
| Volle (2005) | One group only pre-posttest | Elicited |
| Wang (2010) | Pre-postest control (F2F) vs. experiment | Elicited |
| Xiao (2007) | Pre-postest control (F2F) vs. experiment | Naturalistic |
| Yang (2006) | One group only pre-posttest | Naturalistic |
| Zeng (2010) | Posttest only control (F2F) vs. experiment | Naturalistic |

1. **Measures, assessment task and reported reliability in primary studies**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Elicited Language Data | | |  | Naturalistic language Data | | |
|  |  | Language test | |  | Activity/task | Data type | Reliability |
| Study | Performance-based production tasks | Language  test | Standardized  test |  |  |  |  |
| Abrams (2003) | — | — | — |  | Oral discussion | Transcripts | Inter-rater |
| AbuSeileek (2007) | Oral interview | — | — |  | — | — | Inter-rater |
| Ahn (2006) | Oral interview | — | — |  | — | — | Inter-rater |
| Alastuey (2010) | open-ended two-way information exchange | — | — |  | — | — | Inter-rater |
| Alastuey (2011) | Oral Power Point presentation | Proficiency test | — |  | — | — | Test |
| Blake *et al*. (2008) | — | — | Versant for Spanish |  | — | — | Test |
| Blake (2009) | Audio recordings (open-ended response to a written prompt ) | — | — |  | — | — | N.A. |
| Chang (2007) | — | — | — |  | Teacher-student conversation on a topic | Transcripts | N.A. |
| Chang (2008) | Prepared speech | — | — |  | — | — | Test |
| Chen (2008) | — | — | GEPT |  | — | — | Test |
| Huang & Hung (2010) | Audio recordings (voice their opinions on a topic) | — | — |  | — | — | Inter-rater |
| Kost (2004) | — | — | ACTFL |  | — | — | Inter-rater |
| Li (2008) | Recorded interview comprised of response to five open-ended questions | — | — |  | — | — | Inter-rater |
| Lord (2008) | Audio recordings: reading aloud texts | — | — |  | — | — | Inter-rater |
| Payne &Whitney (2002) | Audio recordings: response to a topic | — | — |  | — | — | Inter-rater |
| Pyun (2003) | — | — | — |  | Two 15-minute discussion on topics | Transcripts | N.A. |
| Sanders (2005) | Oral interview (OPI) via telephone | — | — |  | — | — | N.A. |
| Satar& Özdener (2008) | Multiple (picture descriptions, responses to questions, role play, discussions) | — | — |  | — | — | Inter-rater |
| Sequeira (2009) | — | — | ACTFL |  | — | — | Inter-rater |
| Sun (2012) | — | — | — |  | Voice-blog entries | Transcripts | Inter-rater |
| Volle (2005) | Voice email and oral conversation | — | — |  | — | — | Inter-rater |
| Wang (2010) | speech giving | — | — |  | — | — | Inter-rater |
| Xiao (2007) | — | — | — |  | Videoconferencing sessions | Transcripts | Inter-rater |
| Yang (2006) | — | — | — |  | Written discussion | Transcripts | Inter-rater |
| Zheng (2010) | — | — | — |  | Informal debates, discussion of issues after video watching | Transcripts | Inter-rater |

Versant for Spanish: read aloud, listen and repeat, say the opposite, answer short questions, build sentences from jumbled-up word combinations, answer open-ended

GEPT: read aloud, repeat, Q and A

1. **Coding of study characteristics**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Study | N | Setting | Interlocutor | Temporalitya | Mode | Durationb | Grouping | Tool | Publicationc | L2 | L1 |
| Abrams (2003) | 96 | EFL | Peers | B | Both | S (13 weeks) | Mixed | WebCT | J | German | Mixed |
| AbuSeileek (2007) | 131 | EFL | Peers | B | Voice | L (16 weeks) | Mixed | Forum | J | English | Arabic |
| Ahn (2006) | 16 | ESL | Peers | S | Voice | L (1 semester) | Small | Skype | D | English | Mixed |
| Alastuey (2010) | 70 | NA | Peers | S | Voice | S (45 hours/15 weeks) | Pair | Skype | J | English | Mixed |
| Alastuey (2011) | 49 | EFL | Peers | S | Voice | S (15 weeks) | Small | Skype | J | English | Mixed |
| Blake *et al*. (2008) | 334 | ESL | Peers | S | Text | S (10 weeks) | NA | WebCT | J | Spanish | Mixed |
| Blake (2009) | 48 | ESL | Peers | S | Text | S (6 weeks ) | NA | Adobe Breeze | J | English | Mixed |
| Chang (2007) | 50 | EFL | Teacher | S | Voice | S (12 minutes of SCMC;  3 minutes of F2F) | Class | MSN | D | English | Chinese |
| Chang (2008) | 104 | EFL | Peers | S | Text | L (18 weeks) | Pair | Chatroom | J | English | Chinese |
| Chen (2008) | 113 | EFL | Peers | B | Voice | S (8 weeks) | Pair | Mixed | T | English | Chinese |
| Huang & Hung (2010) | 30 | EFL | Peers | A | Voice | L (one semester) | Class | Blog | J | English | Chinese |
| Kost (2004) | 94 | EFL | Peers | S | Voice | L (one semester) | Small | IRC | D | German | Mixed |
| Li (2008) | 90 | EFL | Peers | S | Voice | L (16 weeks) | Small | Chatroom | T | English | Chinese |
| Lord (2008) | 16 | NA | Peers | A | Voice | L (one semester) | Small | Podcast | J | Spanish | NA |
| Payne & Whitney (2002) | 58 | Mixed | Peers | S | Text | S (15 weeks) | Small | Chatroom | J | Spanish | Mixed |
| Pyun (2003) | 20 | Mixed | NA | S | Text | S (an academic quarter) | Pair | MSN | D | English | Korean |
| Sanders (2005) | 55 | ESL | Peers | B | Text | L (2 years) | NA | WebCT | J | Spanish | Mixed |
| Satar & Özdener (2008) | 90 | EFL | Peers | S | Both | S (40~45 minute\*4 sessions(160-180 minutes/4 weeks) | Pair | RD platform | J | English | Turkish |
| Sequeira (2009) | 58 | EFL | Peers | S | Text | L (one semester/4 months) | Pair | Moodle | D | Spanish | English |
| Sun (2012) | 46 | EFL | Peers | A | Voice | L (one semester) | Pair | Blog | J | English | Chinese |
| Volle (2005) | 19 | EFL | Peers | B | Voice | L (one semester) | Mixed | Email | J | Spanish | English |
| Wang (2010) | 52 | EFL | Peers | S | Voice | S (14 weeks) | Small | Chatroom | T | English | Chinese |
| Xiao (2007) | 20 | EFL | Mixed | S | Voice | S (10 weeks) | Pair | Skype | D | English | Chinese |
| Yang (2006) | 33 | EFL | Peers | S | Text | L (16 weeks) | Large | MSN | D | English | Chinese |
| Zheng (2010) | 20 | EFL | Peers | S | Text | L (one semester) | Mixed | Chatroom | T | English | Chinese |

*Note.* a B: both synchronous and asynchronous; A: asynchronous; S: synchronous; b S: Short <=15 weeks; L: Long >15 weeks; c J: Journal article; T: Thesis; D: Dissertation.

1. **Meta-analytic data for the 25 included studies**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Na | Kb | Hedges’s g | S.E. | Lower Upper | | p-Value |
| Abrams (2003) | 96 | 4 | -0.19 | 0.34 | -0.86 | 0.48 | 0.57 |
| AbuSeileek (2007) | 131 | 1 | 1.35 | 0.19 | 0.98 | 1.73 | 0.00 |
| Ahn (2006) | 16 | 1 | 0.12 | 0.34 | -0.56 | 0.79 | 0.74 |
| Alastuey (2010) | 70 | 1 | 0.52 | 0.22 | 0.09 | 0.96 | 0.02 |
| Alastuey (2011) | 49 | 1 | 0.30 | 0.40 | -0.49 | 1.09 | 0.45 |
| Blake (2009) | 334 | 1 | 0.25 | 0.25 | -0.23 | 0.74 | 0.30 |
| Blake *et al*. (2008) | 48 | 1 | -0.06 | 0.21 | -0.47 | 0.34 | 0.77 |
| Chang (2007) | 104 | 1 | -0.46 | 0.31 | -1.06 | 0.15 | 0.14 |
| Chang (2008) | 50 | 2 | 0.29 | 0.20 | -0.10 | 0.67 | 0.14 |
| Chen (2008) | 113 | 1 | 1.49 | 0.44 | 0.62 | 2.36 | 0.00 |
| Huang & Hung (2010) | 30 | 3 | 0.62 | 0.36 | -0.10 | 1.33 | 0.09 |
| Kost (2004) | 94 | 1 | -0.03 | 0.18 | -0.37 | 0.31 | 0.87 |
| Li (2008) | 90 | 1 | 0.73 | 0.24 | 0.25 | 1.21 | 0.00 |
| Lord (2008) | 16 | 1 | 0.94 | 0.36 | 0.22 | 1.65 | 0.01 |
| Payne & Whitney (2002) | 58 | 1 | 0.33 | 0.20 | -0.06 | 0.71 | 0.10 |
| Pyun (2003) | 20 | 2 | -0.04 | 0.39 | -0.81 | 0.72 | 0.91 |
| Sanders (2005) | 55 | 1 | -0.29 | 0.26 | -0.80 | 0.23 | 0.28 |
| Satar & Ozdene (2008) | 90 | 1 | 1.61 | 0.21 | 1.20 | 2.01 | 0.00 |
| Sequeira (2009) | 58 | 1 | 0.85 | 0.28 | 0.31 | 1.39 | 0.00 |
| Sun (2012) | 46 | 4 | -0.28 | 0.13 | -0.53 | -0.04 | 0.02 |
| Volle (2005) | 19 | 3 | 0.29 | 0.28 | -0.26 | 0.85 | 0.30 |
| Wang (2010) | 52 | 5 | 0.23 | 0.31 | -0.37 | 0.84 | 0.45 |
| Xiao (2007) | 20 | 3 | 1.82 | 0.52 | 0.81 | 2.84 | 0.00 |
| Yang (2006) | 33 | 4 | 0.12 | 0.24 | -0.36 | 0.60 | 0.62 |
| Zheng (2010) | 20 | 4 | -0.62 | 0.96 | -2.51 | 1.27 | 0.52 |
| Summary: | 1,712 | 49 | 0.40 | 0.13 | 0.15 | 0.65 | 0.002 |

Note:Na indicates total sample size for each study, and Kb indicates the number of effect sizes contributed by each study.

1. **Average effect sizes for naturalistic and elicited data conditions**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Assessment task | N ES S.E. Upper Lower | | | | |
| Naturalistic | Interaction Transcripts | 7 | -0.15 | 0.10 | -0.34 | 0.04 |
|  | Average | 7 | -0.15 | 0.10 | -0.34 | 0.04 |
| Elicited | Speech-giving | 2 | 0.27 | 0.17 | -0.05 | 0.60 |
|  | Oral Power-point presentation | 1 | 0.30 | 0.40 | -0.49 | 1.09 |
|  | Oral interview | 5 | 0.46 | 0.10 | 0.25 | 0.66 |
|  | Read aloud | 1 | 0.94 | 0.36 | 0.22 | 1.65 |
|  | Response to topics | 4 | 0.44 | 0.12 | 0.20 | 0.68 |
|  | Information exchange | 1 | 0.52 | 0.22 | 0.09 | 0.96 |
|  | Average | 18 | 0.50 | 0.06 | 0.39 | 0.62 |

Q(7)=42.021, p=.000

1. **Effect sizes for oral proficiency traits**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Oral Component | Study | Hedge’s g | S.E. | Lower Upper | | p-Value |
| Accuracy  (k=7) | Chang (2007) | -0.98 | 0.30 | -1.56 | -0.40 | 0.00 |
| Pyun (2003) | 0.81 | 0.34 | 0.14 | 1.48 | 0.02 |
| Sun (2012) | -0.30 | 0.13 | -0.55 | -0.06 | 0.02 |
| Volle (2005) | -0.01 | 0.32 | -0.63 | 0.61 | 0.98 |
| Xiao (2007) | 2.07 | 0.54 | 1.01 | 3.12 | 0.00 |
| Wang (2010) | -1.63 | 0.32 | -2.25 | -1.01 | 0.00 |
| Zheng (2010) | -0.97 | 0.76 | -2.47 | 0.52 | 0.20 |
| Mean ES |  | -0.17 | 0.35 | -0.86 | 0.52 | 0.63 |
| Fluency  (k=7) | Abrams (2003) | -0.24 | 0.61 | -1.44 | 0.96 | 0.70 |
| Huang & Hung (2010) | 0.66 | 0.37 | -0.05 | 1.38 | 0.07 |
| Sun (2012) | -0.35 | 0.13 | -0.60 | -0.11 | 0.01 |
| Xiao (2007) | 1.78 | 0.51 | 0.78 | 2.78 | 0.00 |
| Yang (2006) | 0.01 | 0.24 | -0.46 | 0.49 | 0.95 |
| Wang (2010) | -1.88 | 0.23 | -2.33 | -1.42 | 0.00 |
| Zheng (2010) | -1.32 | 1.72 | -4.70 | 2.06 | 0.44 |
| Mean ES |  | -0.11 | 0.41 | -0.91 | 0.68 | 0.78 |
| Lexical level  (k=7) | Wang (2010) | 1.71 | 0.32 | 1.08 | 2.34 | 0.00 |
| Zheng (2010) | 0.61 | 0.33 | -0.03 | 1.24 | 0.06 |
|  | Abrams (2003) | -0.29 | 0.18 | -0.63 | 0.05 | 0.10 |
| Abrams (2003) | -0.27 | 0.18 | -0.61 | 0.08 | 0.13 |
| Yang (2006) | -0.01 | 0.24 | -0.49 | 0.46 | 0.96 |
|  | Yang (2006) | 0.13 | 0.24 | -0.34 | 0.61 | 0.58 |
| Huang & Hung (2010) | 0.81 | 0.37 | 0.08 | 1.54 | 0.03 |
| Mean ES |  | 0.34 | 0.24 | -0.14 | 0.81 | 0.17 |
| Pronunciation  (k=4) | Alastuey (2010) | 0.53 | 0.22 | 0.09 | 0.96 | 0.02 |
| Lord (2008) | 0.94 | 0.36 | 0.22 | 1.65 | 0.01 |
| Sun (2012) | -0.21 | 0.12 | -0.45 | 0.04 | 0.09 |
| Volle (2005) | 0.27 | 0.18 | -0.09 | 0.63 | 0.15 |
| Mean ES |  | 0.31 | 0.23 | -0.15 | 0.77 | 0.18 |
| Syntactic level  (k=8) | Abrams (2003) | 0.02 | 0.18 | -0.34 | 0.38 | 0.91 |
| Huang & Hung (2010) | 0.38 | 0.36 | -0.33 | 1.08 | 0.29 |
| Pyun (2003) | -0.90 | 0.44 | -1.75 | -0.04 | 0.04 |
| Sun (2012) | -0.27 | 0.12 | -0.52 | -0.03 | 0.03 |
| Xiao (2007) | 1.63 | 0.50 | 0.65 | 2.61 | 0.00 |
| Yang (2006) | 0.35 | 0.25 | -0.13 | 0.83 | 0.15 |
| Wang (2010) | 2.51 | 0.37 | 1.79 | 3.23 | 0.00 |
| Zheng (2010) | -0.78 | 0.23 | -1.23 | -0.34 | 0.00 |
| Mean ES |  | 0.32 | 0.30 | -0.27 | 0.91 | 0.29 |
| Holistic  (k=16) | AbuSeileek (2007) | 1.36 | 0.19 | 0.98 | 1.73 | 0.00 |
| Ahn (2006) | 0.12 | 0.34 | -0.56 | 0.79 | 0.74 |
| Alastuey (2011) | 0.30 | 0.40 | -0.49 | 1.09 | 0.45 |
| Blake (2009) | 0.25 | 0.25 | -0.23 | 0.74 | 0.30 |
| Blake *et al*. (2008) | -0.06 | 0.21 | -0.47 | 0.34 | 0.77 |
| Chang (2008) | 0.29 | 0.20 | -0.10 | 0.67 | 0.14 |
| Chang (2007) | 0.07 | 0.32 | -0.57 | 0.70 | 0.84 |
| Chen (2008) | 1.49 | 0.44 | 0.62 | 2.36 | 0.00 |
| Kost (2004) | -0.03 | 0.18 | -0.37 | 0.31 | 0.87 |
| Payne & Whitney (2002) | 0.33 | 0.20 | -0.06 | 0.71 | 0.10 |
| Sanders (2005) | -0.29 | 0.26 | -0.80 | 0.23 | 0.28 |
| Satar & Özdener (2008) | 1.61 | 0.21 | 1.20 | 2.01 | 0.00 |
| Sequeira (2009) | 0.85 | 0.28 | 0.31 | 1.39 | 0.00 |
| Volle (2005) | 0.61 | 0.33 | -0.02 | 1.25 | 0.06 |
| Wang (2010) | 0.46 | 0.28 | -0.09 | 1.01 | 0.10 |
| Li (2008) | 0.73 | 0.24 | 0.25 | 1.21 | 0.00 |
| Mean ES |  | 0.50 | 0.15 | 0.20 | 0.79 | 0.00 |
|  |  |  |  |  |  |  |

Q(5)=4.458, p=.486

1. **Effect sizes for studies with different treatment durations**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Duration** | **Study** | **ES** | **S.E** | **Lower** | **Upper** |
| LONG | AbuSeileek (2007) | 1.35 | 0.19 | 0.98 | 1.73 |
| Ahn (2006) | 0.12 | 0.34 | -0.56 | 0.79 |
| Chang (2008) | 0.29 | 0.20 | -0.10 | 0.67 |
| Huang & Hung (2010) | 0.62 | 0.36 | -0.10 | 1.33 |
| Kost (2004) | -0.03 | 0.18 | -0.37 | 0.31 |
| Lord (2008) | 0.94 | 0.36 | 0.22 | 1.65 |
| Sanders (2005) | -0.29 | 0.26 | -0.80 | 0.23 |
| Sequeira (2009) | 0.85 | 0.28 | 0.31 | 1.39 |
| Sun 2012 | -0.28 | 0.13 | -0.53 | -0.04 |
| Volle (2005) | 0.29 | 0.28 | -0.26 | 0.85 |
| Yang (2006) | 0.12 | 0.24 | -0.36 | 0.60 |
| Li (2008) | 0.73 | 0.24 | 0.25 | 1.21 |
| Zheng (2010) | -0.62 | 0.96 | -2.51 | 1.27 |
| Average (k=13) | 0.25 | 0.06 | 0.12 | 0.37 |
| SHORT | Abrams (2003) | -0.19 | 0.34 | -0.86 | 0.48 |
| Alastuey (2010) | 0.52 | 0.22 | 0.09 | 0.96 |
| Alastuey (2011) | 0.30 | 0.40 | -0.49 | 1.09 |
| Blake (2009) | 0.25 | 0.25 | -0.23 | 0.74 |
| Blake *et al*. (2008) | -0.06 | 0.21 | -0.47 | 0.34 |
| Chang (2007) | -0.46 | 0.31 | -1.06 | 0.15 |
| Chen (2008) | 1.49 | 0.44 | 0.62 | 2.36 |
| Payne & Whitney (2002) | 0.33 | 0.20 | -0.06 | 0.71 |
| Pyun (2003) | -0.04 | 0.39 | -0.81 | 0.72 |
| Satar & Özdener (2008) | 1.61 | 0.21 | 1.20 | 2.01 |
| Xiao (2007) | 1.82 | 0.52 | 0.81 | 2.84 |
| Wang (2010) | 0.23 | 0.31 | -0.37 | 0.84 |
|  | Average (k=12) | 0.44 | 0.08 | 0.29 | 0.60 |

Q(1)=3.679, p=.055