Measuring Writing Improvement in a Graduate-level Turkish University: WAC Approach, Objective Standardized Measurement, and Results

David R. Albachten

Academic Writing Center and Graduate Writing Program
Istanbul Sehir University, Istanbul, Turkey

Measuring Writing Improvement

Agenda

* Background
* Objective
* Database
* Measurement Tools
* Methodology
* Results
* Implications
* Questions . . . mine and yours
* References

Measuring Writing Improvement

Background (1)

* Istanbul Sehir University -- a new, private foundation-sponsored, English language medium, with BA, BS, MA, MS, and PhD degrees
  * 100% are L2 English speakers
    * 85% Turkish
    * remainder from dozens of countries (Middle East, eastern Europe, Africa, and Asia)
  * All students provided a laptop upon admission
  * All assignments typed and transmitted as Microsoft Word attachments
  * Writing Across the Curriculum (WAC) in all Institute of Social Sciences classes (average of one writing assignment every two-three weeks per class)

Measuring Writing Improvement

Background (2)

* Academic Writing Center (AWC) established for the Institute to serve graduate students (MA) in Winter 2011
  * AWC expanded in 2012 (includes all post-graduates and faculty)
* Critical Academic Writing course (CAW 501) established for the Institute as a mandatory class in Winter 2011
  * students are not allowed to write their dissertation until they pass CAW 501

Measuring Writing Improvement

Objective

* Can we prove -- through the AWC, CAW, or both -- if there is writing improvement?
  * Building an objective, reproducible methodology for measurement
    * grammar, punctuation, usage
    * appropriate documentation (plagiarism avoidance)
    * (**of course** . . . structure, content, flow, and readability are hygiene factors)
  * Measuring students papers at appropriate levels of intervention (at least 4 times across 4 semesters, over two years)
    * pre-AWC, post-AWC (4-8 interventions)
    * pre-CAW, post-CAW (6)
    * pre- and post-CAW <u>and</u> AWC (10-14)
    * controls (neither AWC nor CAW) (zero)

Measuring Writing Improvement

The Graduate Writing Program/Academic Writing Center Electronic Microsoft Word **Database** (total possible pool) (2011-2014 (ytd)):

* Academic Writing Center:
  * n = 200 students
  * 954 submissions
  * 18,230 pages (average 19 pages/submission)
* Critical Academic Writing courses:
  * n = 136 students
  * 723 submissions
  * 1,422 pages (average 2 pages/submission)

* Post-CAW and AWC (both CAW and AWC feedback, n = 52)

* Controls (neither AWC nor CAW, n = 25)

Measuring Writing Improvement

Measurement Tools

* ETS e-rater (v. 11.1)
    * grammar: s/v, run-on, fragment . . .
    * mechanics: punctuation, capitalization . . .
    * style: passive voice, long and short . . .
    * usage: articles, wrong form . . .
    * spelling: spelling (adjusted, minus foreign words)
    * **grammar score = errors per 100 words** (range 5.4– 0.5)
* iThenticate (stand alone v. 2.0.3)
    * similarity score **adjusted similarity index = similarity score minus appropriate use** (33% - 0%)
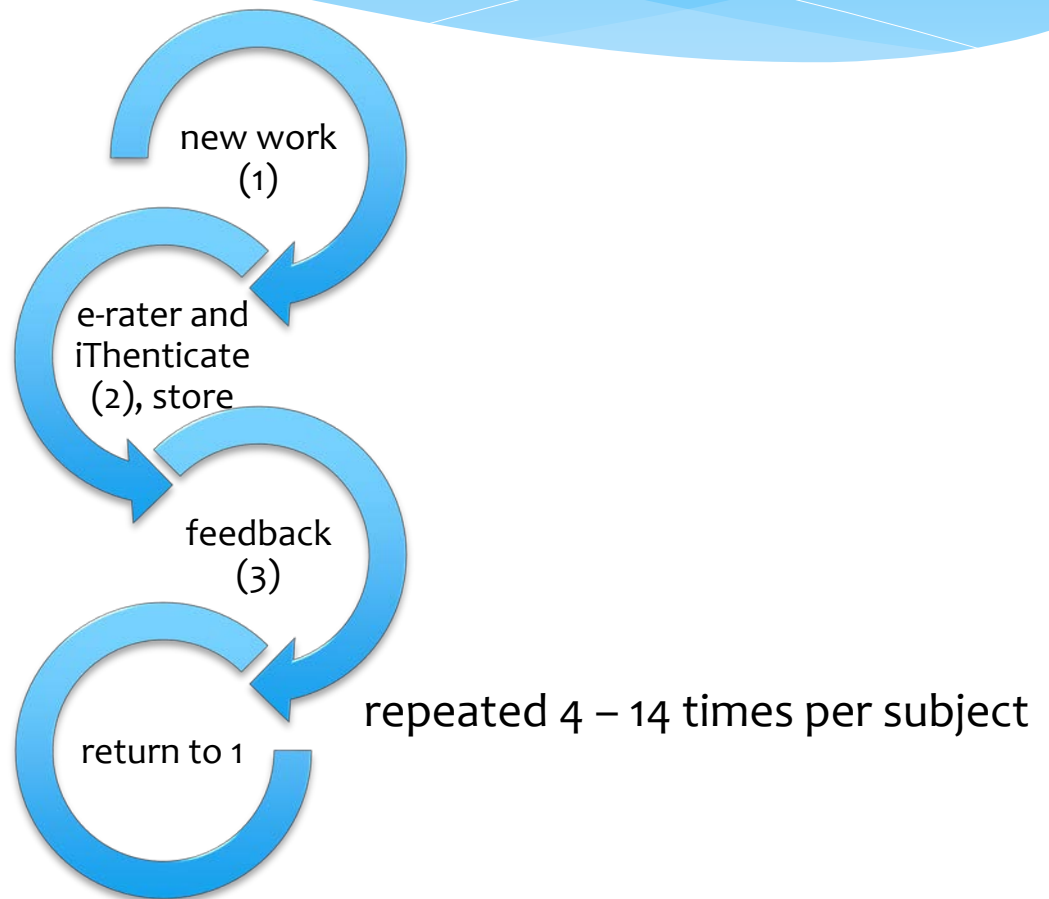
Measuring Writing Improvement

Methodology

* 36 students, 112 separate papers totaling 1,680 pages were subjected to measurement (100% L1 Turkish)
  * 10 controls (316 pages, 26 papers) about a third
  * 26 interventions (1,364 pages, 86 papers) about two thirds
* iThenticate and ETS e-rater run on the papers as they came in (representing <u>first draft </u>student effort), scores set aside
* <u>Normal feedback occurred</u>, AWC and CAW using "Track Changes" comments in Microsoft Word in addition to face-to-face sessions
* Then, <u>new</u> papers (<u>not</u> revised versions, but <u>new</u> student work) from students were again run for subsequent observations (ETS and iThenticate, scores set aside) and feedback again given <u>as usual</u>
  * Each ETS e-rater error count and similarity index were captured for each new paper (42 elements per paper) . . . 8,031 errors tracked in this study
* The papers rated in this study represent a time span of four semesters over two years

Measuring Writing Improvement

Methodology

new work (1)

e-rater and iThenticate (2), store

feedback (3)

return to 1

repeated 4 – 14 times per subject

Measuring Writing Improvement

Results (1) . . . documentation

* iThenticate first visit results:
    * high: 33%
    * low: 0%
    * average: 15.2%
* iThenticate second visit results:
    * high: 10%
    * low: 0%
    * average: 3.2%
* iThenticate result stayed low after second visit
    * did not prove to be an interesting measure, past the first visit

78%
reduced

Measuring Writing Improvement

Results (2) . . . common errors

* Most common ETS e-rater errors (all subjects):
  * article error and missing article (1.37) (more than 5X next closest)
  * missing comma (0.27) (relative clause . . . ", which" and series)
  * spelling (0.25)
  * sentence fragment (0.25)
  * run-on sentence (0.14)
  * subject verb agreement (0.08)
  * passive voice (0.06)
  * wrong preposition (0.04)
  * missing period (0.01)
  * TOTAL: 2.47 errors per 100 words of a total of 3.12 (80-20 rule)

Measuring Writing Improvement

Results (2.1) errors ⭐ = discoverable with MS Word checkers (25%)

* Most common ETS e-rater errors (all subjects):
  * article error and missing article (1.37) (more than 4.5 times next closest)
  * missing comma (0.27) (relative clause and series) ⭐
  * spelling (0.25) ⭐
  * sentence fragment (0.25)
  * run-on sentence (0.14)
  * subject verb agreement (0.08)
  * passive voice (0.06) ⭐
  * wrong preposition (0.04)
 * missing period (0.01) ⭐

Measuring Writing Improvement

Results (2.2) errors ⭐ = Turkish language issue (40%)

* Most common ETS e-rater errors (all subjects):
  * article error and missing article (1.37) (more than 4.5 times next closest) ⭐
  * missing comma (0.27) (relative clause and series) ⭐
  * spelling (0.25)
  * sentence fragment (0.25)
  * run-on sentence (0.14)
  * subject verb agreement (0.08)
  * passive voice (0.06)
  * wrong preposition (0.04)
  * missing period (0.01)

Measuring Writing Improvement

Results (3) . . . reduction of errors by intervention & control

56%
reduction
AWC

64%
reduction
CAW

79%
reduction
CAW+AWC

3%
increase
control
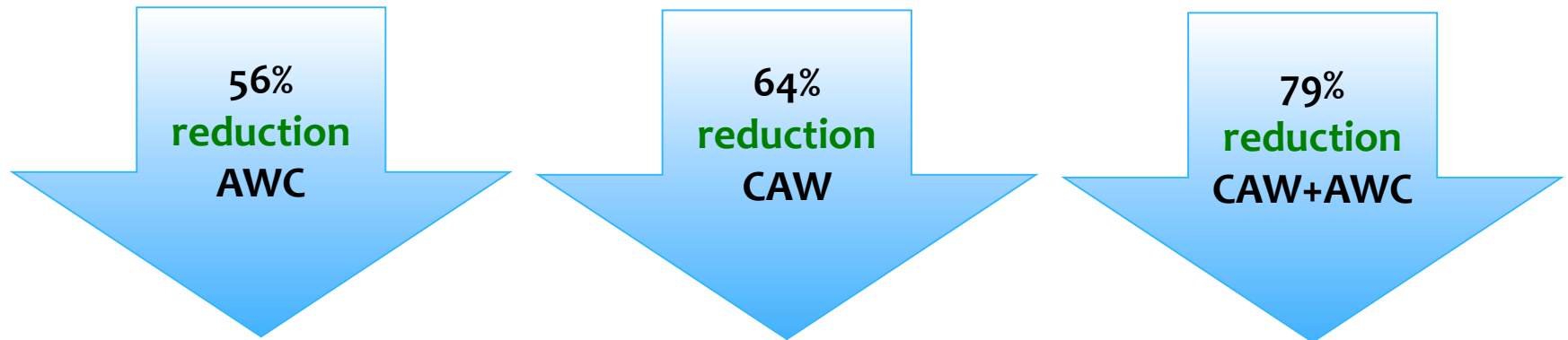
* intervention group (average reduction)
  * AWC
  * CAW
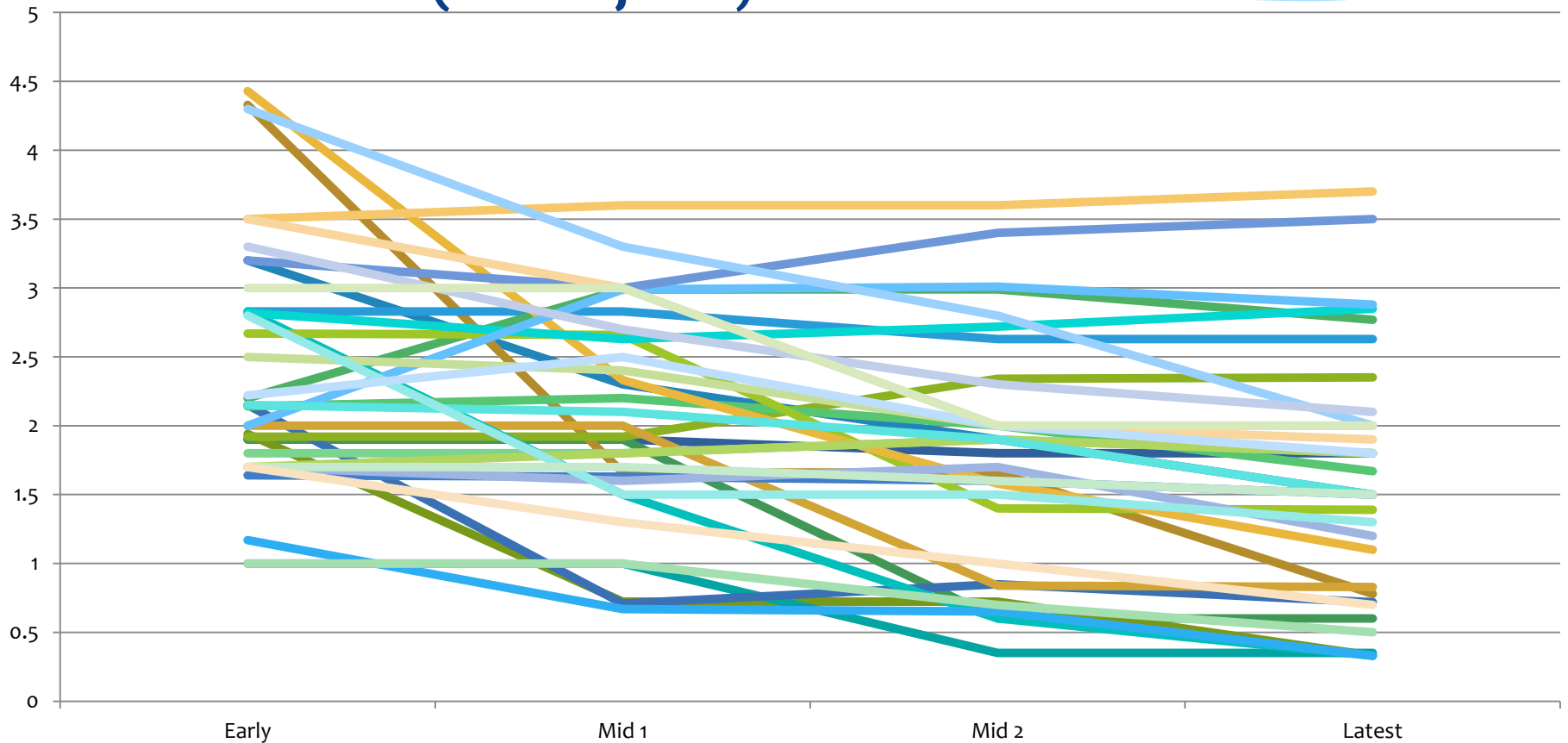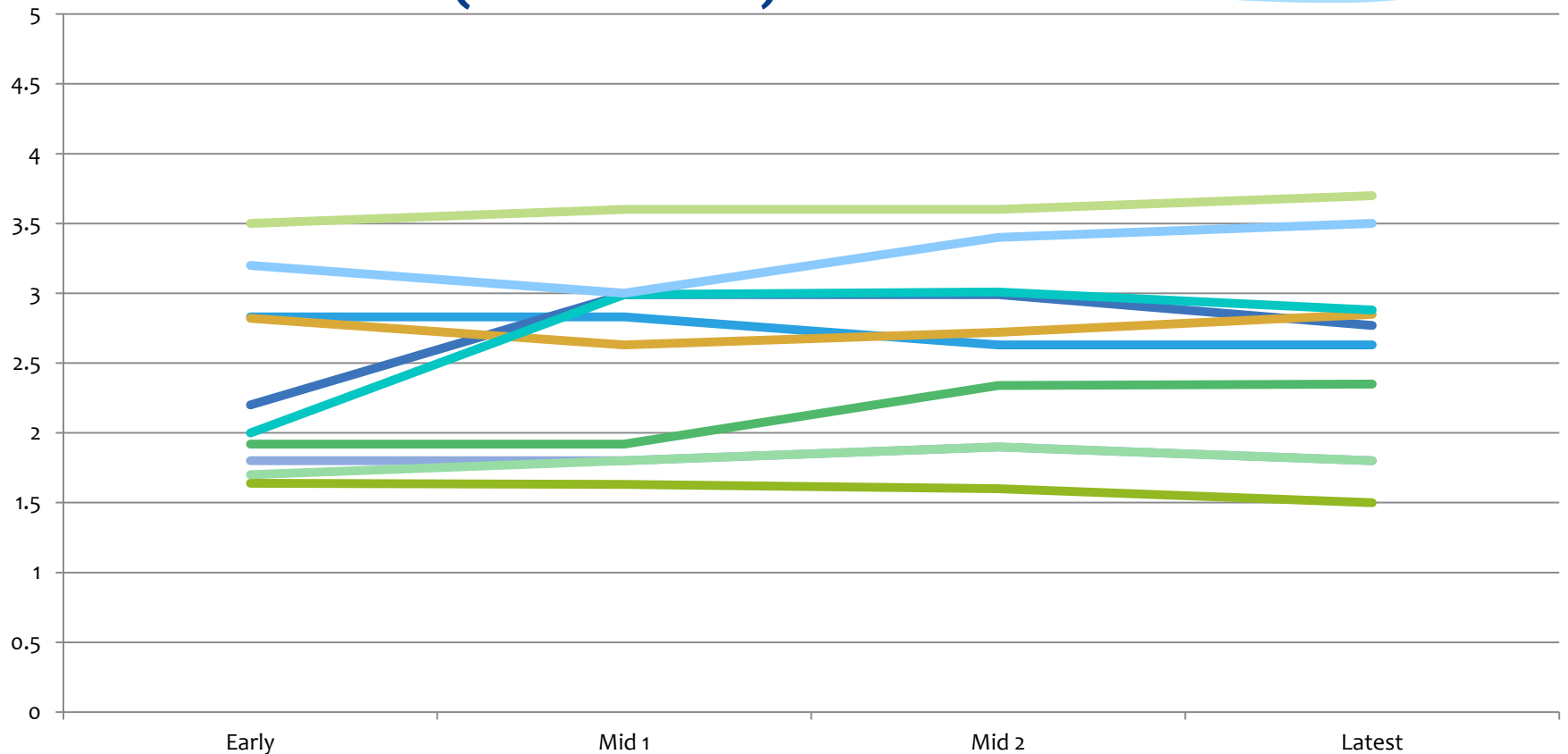 * AWC+CAW
* control group

Measuring Writing Improvement

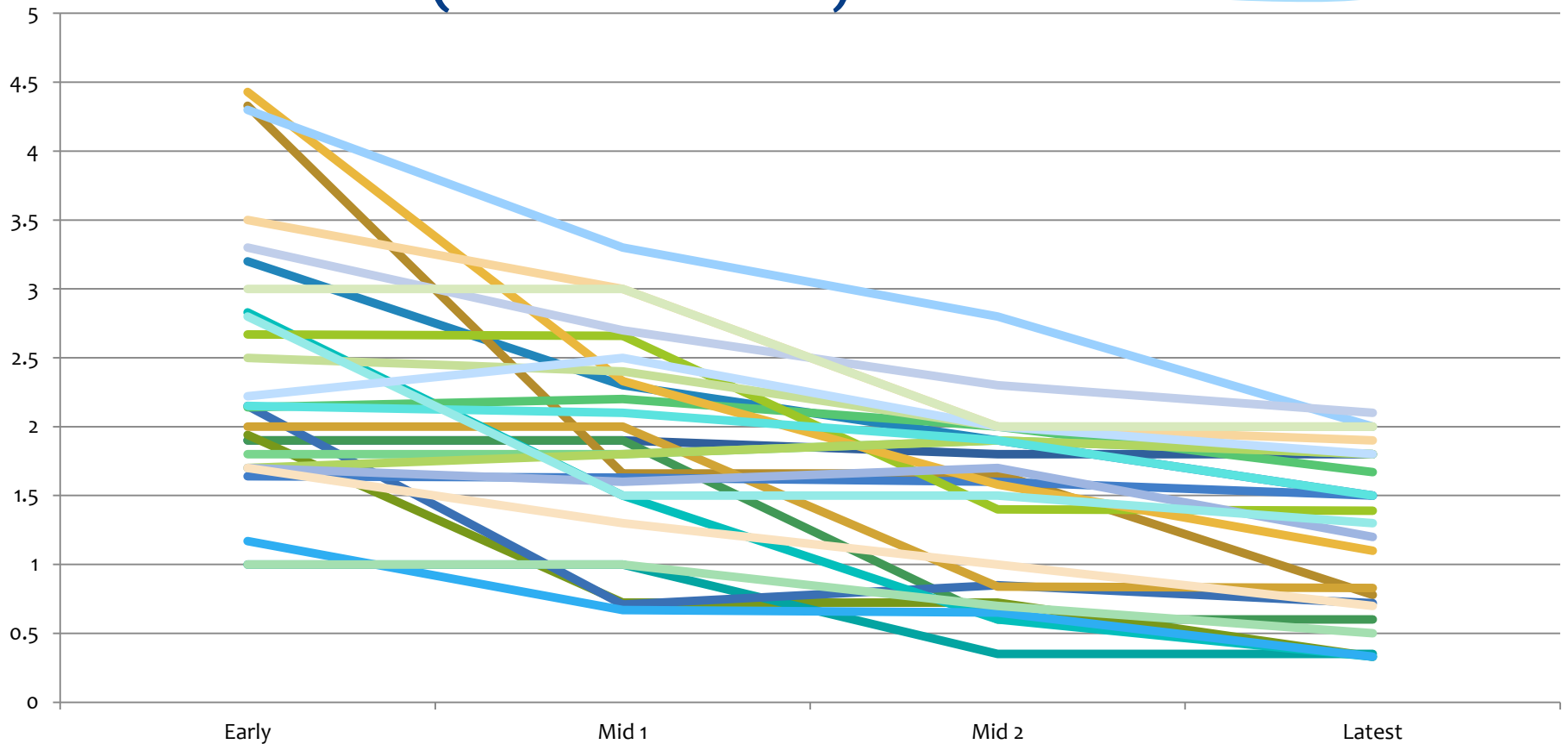**Results: errors (all subjects)**

Measuring Writing Improvement

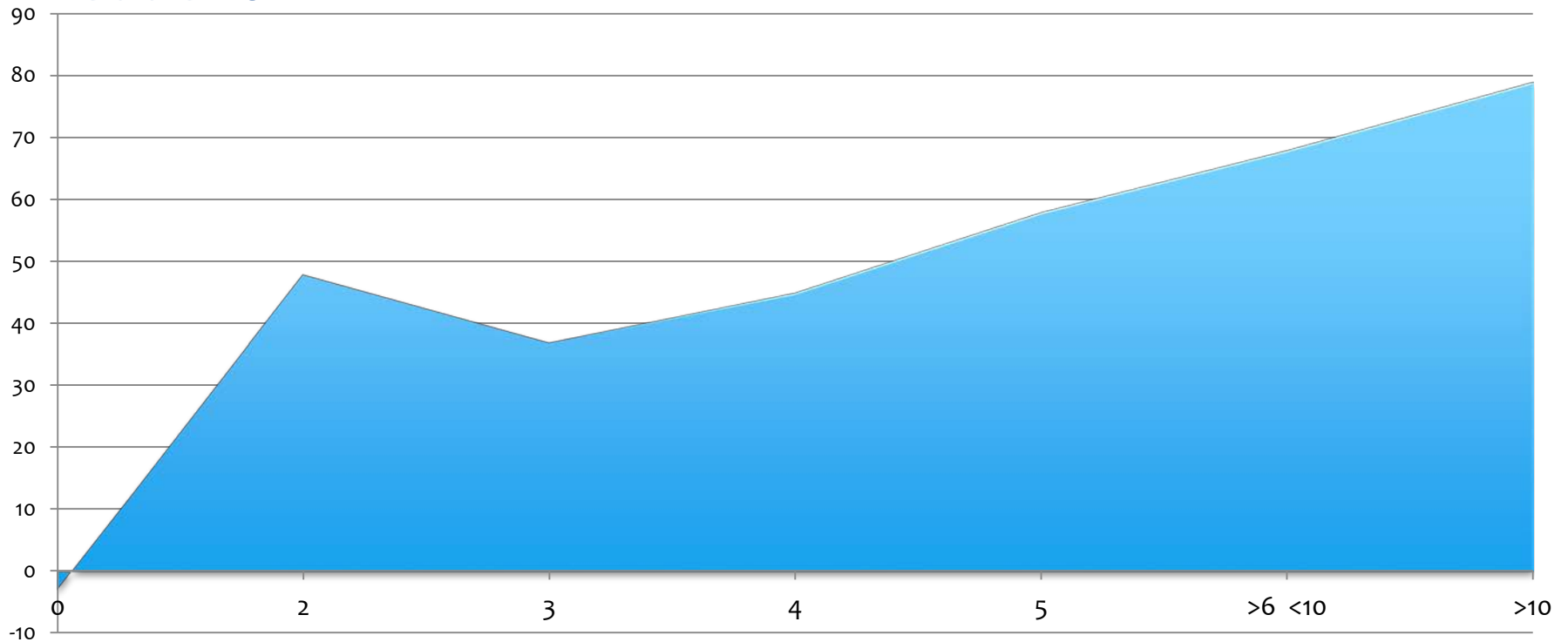**Results: errors (10 controls)**

Measuring Writing Improvement

Results: errors (minus controls)

Measuring Writing Improvement

Results: intervention number vs. improvement in error reduction

Measuring Writing Improvement

Implications

* Proper documentation can be learned quickly
* 9 errors represent 80% of all errors, 3 items represent more than 50% of errors (articles, clauses, spelling)
    * Could serve to focus teaching (particularly Turkish L1)
    * Proper use of MS Word checkers could reduce errors 25%
        * computer training!
* **Writing alone does not improve writing!**
* **Writing with extensive feedback improves writing!!**
    * **additional writing with feedback improves writing more!!!**
        * an upper threshold has not yet been found

Measuring Writing Improvement

Questions . . . mine and yours

* Mine
    * better grammar = better writing?
        * at the graduate level I say, "Yes"
            * grammar issues <u>are</u> higher order issues for these students
        * subject teachers and the students tell me so
    * true for a much larger sample?
        * ask me in a couple of years
    * are there better ways to objectively and reproducibly measure?
* Yours?

Measuring Writing Improvement

References

Burstein, J. (2012) **Fostering Best Practice in Writing Instruction and Assessment with E-rater** in *Writing Assessment in the 21st Century: Essays in Honor of Edward M. White,* N. Elliott and L. Perelman (Eds.), Hampton Press.

Cogie, J., Strain, K. and Lorinskas, S. (1999) **Avoiding the Proofreading Trap: The Value of Error Correction Process,** in *The Writing Center Journal,* V. 19 n. 2, pp. 7 – 32.

McLeod, S. H. and Soven, M. (2000) **Writing Across the Curriculum: A Guide to Developing Programs,** Sage

Madnani, N., Tetreault, J and Chodorow, M. (2012) **Exploring Grammatical Error Correction with Not-So-Crummy Machine Translation,** in *Proceedings of the 7th Workshop on Innovative Use of Natural Language Processing for Building Educational Applications (BEA),* Association for Computational Linguistics

Meyers, S. (2003) **Reassessing The 'Proofreading Trap': ESL Tutoring and Writing Instruction,** in *The Writing Center Journal,* V. 24 n. 1, pp. 51-70.

Williamson, D. M., Xi, X., and Breyer, J. F. (2012) **A Framework for Evaluation and Use of Automatic Scoring,** in *Educational Measurement: Issues and Practices,* Vol. 31, No. 1, pp. 2–13

Xi, X. (2012) **Validity and the Automatic Scoring of Performance Tests,** in *The Routledge Handbook of Language Testing,* pp. 438–451, G. Fulcher & F. Davidson (Eds.), Routledge