DOCUMENT RESUME

ED 041 492 FL 001 219

AUTHOR Bockman, John F.; Valencia, Felizardo L.
TITLE An Experiment in the Development of Conversational Fluency in Spanish Through Use of Electric Cue Board and Related Methodology.
INSTITUTION Tucson Public Schools, Ariz.
PUB DATE Oct 68
NOTE 24p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.30
DESCRIPTORS *Audiolingual Methods, Audiolingual Skills, Conversational Language Courses, Dialogue, Educational Experiments, Educational Innovation, *Electromechanical Aids, Electronic Equipment, Instructional Improvement, Instructional Materials, Language Fluency, *Language Instruction, Modern Languages, Oral Communication, *Secondary Schools, *Spanish, Standard Spoken Usage, Student Evaluation

ABSTRACT

An innovative approach to language instruction, developed by Felizardo Valencia using "Juan, Maria y Pepe" as a basic text, is contrasted with more traditional audiolingual methodology in a first-year Spanish course. A control group of high school students is instructed with the use of a teacher-operated electric cue board with abbreviated cues which allows for controlled teaching of the structures, patterns of communication, tenses, and topics for conversation in a dialogue. Description of the project design, matched-pairs of subjects, results of the study, and group test results are included. Computer generated statistical tables are used in the analysis of the results of the experiment. (RL)

TUCSON PUBLIC SCHOOLS
COORDINATOR OF FOREIGN LANGUAGE INSTRUCTION

AN EXPERIMENT IN THE DEVELOPMENT OF CONVERSATIONAL FLUENCY IN SPANISH THROUGH USE OF ELECTRIC CUE BOARD AND RELATED METHODOLOGY, by John F. Bockman and Felizardo L. Valencia

Contents — Page

Purpose of the Study..... 1

Background of the Study..... 1

The Project Design..... 3

The Matched-Pairs..... 4

Statement of Goals and Objectives of the Experimental Program, Felizardo L. Valencia..... 5

Results of the Study..... 7

Final Conclusions..... 11

Recommendations..... 11

Group Test Results..... 12

Notes..... 22

October, 1968

AN EXPERIMENT IN THE DEVELOPMENT OF CONVERATIONAL FLUENCY IN SPANISH THROUGH USE OF ELECTRIC CUE BOARD AND RELATED METHODOLOGY, by John F. Bockman and Felizardo L. Valencia

The Purpose

The purpose of the study is to compare, in various Spanish language skills, first-year students of Spanish taught by the method developed by Felizardo Valencia of Rincon High School with those taught by several other methods in current use at Rincon, Catalina, and Palo Verde High Schools.

Background of the Study

Felizardo Valencia's method, Juan, María y Pepe, Español Práctico, is an audiolingual one. It has as a primary objective the development of conversational fluency. Mr. Valencia, however, is in fundamental disagreement with many of the assumptions and much of the methodology of the commercial audiolingual programs (e.g. A-LM, Entender y Hablar, Primera Vista).

Mr. Valencia holds that the dialog-narrative approach of the published audiolingual programs permits, even encourages, less than is necessary and less than can be done to achieve control of topic and structure with native-like speed or fluency. The Valencian method, therefore, eschews the dialog-narrative approach. It employs a system of controlled dialogs of maximum flexibility and of maximum relevance and vital concern to the student.

To achieve control of topic and structure with native-like speed or fluency, the teacher uses an ingenious electric cue board which includes abbreviated cues that control all possible structures, patterns of communication, tenses, and topics for conversation. The teacher controls the cue board through a portable console with a switch for each cue. As the lights turn on at lightning-fast speed, the students respond and thereby are

conditioned to a rote skill. The cueing system is not simply a device. It is an analogue of Mr. Valencia's analysis of communicative relationship and of patterns, structures, verbal elements, etc. The cueing system is, of course, essential to the method.

Mr. Valencia had experimented with the method, but without the aid of the cue board, during adult evening Spanish classes and with second-year students of Spanish at Rincon High School in 1966-67 who were otherwise using Segunda Vista. Subsequently, Mr. Valencia requested and was granted two experimental first-year Spanish classes at Rincon for the 1967-68 school year. A cue board was built to his specifications and installed in room 204, Rincon High School. At present the experiment continues into the second year with those students who completed the first year of the program.

Assuming that valid measuring instruments were available, it was thought that the experiment should eventually lead to a determination of whether students trained by this method could indeed be brought to score higher on speaking and fluency tests than do students trained by the other methods. It would be of interest also to determine how they compare in pronunciation and in listening comprehension. With the heavy emphasis on achieving an optimum degree of spontaneity and fluency, it was considered important to determine how the experimental students compare in the development of reading and writing skills normally achieved by the end of the first year.

Four men teachers, all native or near-native speakers of Spanish, with a reputation for rapport and effectiveness with students, agreed to the participation of seven different classes, two at Catalina, four at Palo Verde, and one at Rincon, as the control group--a total of 147 students. The experimental group consisted of 40 Rincon High School students who had been assigned to Spanish 1-2S by normal, unselective scheduling procedures.

The Pimsleur Language Aptitude Battery was administered to the 187 subjects of this study in September, 1967. Data concerning bilingualism, if any, and prior foreign language study experience were collected at the same time. The Pimsleur Spanish Proficiency Tests in listening comprehension and reading were given to all 187 students in May, 1968. The Pimsleur Spanish Proficiency Tests in speaking and writing were given to 50 students of 25 matched pairs (q.v.).

The Project Design

The study makes two kinds of comparisons: 1) of individuals in matched-pairs of experimental and control students. 2) of the separate teacher-groups.

Pre-test data collected include: sex, month and year of birth, history of bilingualism or previous formal language study if any, grade, aptitude test raw scores: motivation, grade point average, vocabulary, language analysis, sound discrimination, sound-symbol association, total parts 1 through 6, total parts 3 and 4, total parts 5 and 6, verbal aptitude stanine, auditory aptitude stanine, and total aptitude stanine.

Post-test data include: listening proficiency[1] raw score and stanine, reading proficiency raw score and stanine, and final grade for all 187 subjects of the experiment. In addition, post-test data for the 50 students of the 25 matched-pairs include: writing proficiency raw score and stanine, speaking proficiency raw score and rating, fluency raw score, and pronunciation raw score.

From the data punched on IBM cards a variety of correlational studies

---

[1] "The achievement test tests what the student has been taught whereas the proficiency test tests how far he has advanced toward mastery of a language, by whatever method he has been taught."
--Pimsleur

becomes possible, for example, the relationship between total aptitude stanines and listening comprehension proficiency stanines. It is assumed that any significant correlations discovered might permit tentative judgment concerning methodological effectiveness. To assist such interpretation of possibly significant differences, teacher participants in the experiment were asked to submit a statement of their instructional objectives.

The Matched-Pairs

Twenty-five experimental students were matched one-for-one with twenty-five control students by the following variables: sex, ability (IQ) stanine score, grade in school, verbal aptitude raw score and stanine score, and auditory raw score and stanine score. Students from homes with a record of bilingualism and those who had studied Spanish or any other foreign language previously were not included among the matched-pairs.

Among the 25 control students, five were students of teacher C-1 at Catalina, nine of teacher C-2 at Palo Verde, seven of teacher C-3 at Palo Verde, and four of teacher C-4 at Rincon.

With matched-pairs, it is possible not merely to study average group differences but also to analyze the average difference between pairs. The introduction of controls through matching reduces the standard error of difference. This term, which is computed from the standard error of the mean, (a measure of the probable extent to which a mean is apt to vary on future samplings), is used to determine whether any difference found between means is a real or a chance difference. Every measurement has some unreliability. One can never be certain that one particular administration of a test gives the "true" mean score of the group tested. If the given group could take the same test an infinite number of times, each time as though they were taking it the first time, a series of resulting means would distribute themselves in a normal curve, yielding a "true" mean. But since it is impossible to discover

the true mean in this fashion, conclusions can be drawn concerning probable results of measuring broader samplings from the manner in which scores are dispersed around the mean in any sample testing.[2]

Statement of Goals and Objectives of the Experimental Program

Goals

(Primary) To develop, from the very beginning of the program, the ability to speak and converse in depth with the fluency and versatility of a native speaker of Spanish within the limitations imposed by the extent of linguistic matter assimilated at each stage of development.

(Secondary) To develop listening comprehension, reading, and writing skills within the framework of the primary goal.

Objectives

Depending on the stage of development, the student will be able to discuss individually, or as a member of a group, or as a partner in a two- or several-sided conversation with native-like fluency and versatility any of a number of topics of immediate relevance to the student. The topics will be selected jointly by the teacher and the students and will constitute the subject matter of the program.

Student performance will be measured by

a) evaluating his responses to dialog control cues flashed electrically by an investigator using the Electric Cue Board;

b) by observing his participation in natural classroom speaking activities, spontaneous or controlled;

---

[2] Ideas expressed in this paragraph were garnered from Understanding Educational Research, Van Dalen; A Primer of Statistics, Franzblau; Statistical Concepts, Amos; and A Psycholinguistic Experiment in Foreign Language Teaching, Scherer. (See notes.)

c) by teacher-made speaking-fluency tests;

d) by standardized speaking-fluency tests.

The student will be able to read and write dialogs, stories, expositions, etc. constructed of the linguistic matter (as above). Performance will be measured by oral reading, reading comprehension exercises, and by teacher-made and standardized reading and writing tests.

Definition of Terms and Limitations

Because of the innovative nature of this approach, a special definition is given to the following dialog controls used in the program:

a) Pattern of Communication - the factors that determine the direction and number of a given dialog;

b) Structure - those factors which make a communication interrogatory, commanding, expository, or exclamatory, and the cultural factors that might have linguistic effects on the communication;

c) Time - tense applied to verbs;

d) Topic - the subject matter of the communication as well as the vocabulary and other lexical matter to be used in the dialogs. These will be drawn from topics of maximum relevance to ninth through twelfth graders of Rincon High School, Tucson, Arizona. Because of the in-depth treatment of each topic, and behavioral outcomes sought in the students with respect to these situational topics, the vocabulary, for example, will not be as varied and extensive as that learned in regular Spanish classes. There may be many additions.

(A complete and detailed course-of-study of the program which outlines each step of instruction, subject matter, and behavioral objectives is available upon request.)

Results of the Study

Matched-Pairs

Pre-Test Data

There were 25 pairs of matched students in the experiment. Matching was by Ability Stanine, Verbal Language Aptitude Stanine, and Auditory Language Aptitude Stanine, and by sex and age to secure as exact an equation as possible. The mean stanines of the two groups in pre-test data are shown in Table I.

TABLE I
MEAN PRE-TEST STANINES

	X-Group (N-25)	C-Group (N-25)
Ability Stanine (IQ)	6.72	6.56
Verbal Language Aptitude Stanine	5.32	5.16
Auditory Language Aptitude Stanine	6.80	6.84
Total Language Aptitude Stanine	6.44	6.28

Observations

The Total Language Aptitude Stanine includes measurement of Grade Point Average and Motivation as well as measurements of Verbal and Auditory Language Aptitude. The mean Total Aptitude Stanine and mean Ability Stanine of the two groups are very close. The means of Auditory Aptitude are more than one stanine higher than the means of Verbal Aptitude. This discrepancy is a phenomenon noted in language aptitude testing results collected from all groups so far measured in Tucson Public Schools.

Post-Test Data

Standardized Spanish proficiency testing data provide the opportunity to compare the two groups of this study with one another and with the norms group. The norms group comprised 5,306 first-level Spanish students from 89 schools in 34 systems distributed throughout 20 states in the nation (1966). Mean stanines of proficiency in the various Spanish language skills are shown in Table II.

TABLE II

MEAN POST-TEST STANINES

	Norms Group (N-5306)	X-Group (N-25)	C-Group (N-25)
Listening	5.00	4.40	5.28
Reading	5.00	4.80	4.80
Writing	5.00	4.24	5.24*

Results

The t-tests[3] for correlated means reveal:

a) no significant difference in the experimental and control groups in Listening and Reading;

b) the control group is significantly higher in Writing. (*Level of Significance: $p < .01$)[4]

Observations

The mean Writing Stanine of the control group matches the mean Verbal Aptitude Stanine of the group (v. Table I). This may suggest that the teachers

---

[3] The t-test is used to determine whether a difference observed between means should be attributed to chance or whether it is reasonable to conclude that the programs had differential effects on the learning tasks. (See notes.)

[4] I.e., there is one chance in 100 of this being a chance difference. (See Notes.)

of the control group have strong, persistent Writing objectives.

The auditory Aptitude Stanines of the groups might lead to the expectation of higher achievement in Listening. It may be that purposeful Listening per se as a distinct set of objectives is weak in the control programs and weaker in the experimental program. Listening may be incidental to speaking, but may be receiving a somewhat clearer separate identification in the control programs.

Mean Speaking Scores

The Pimsleur Spanish Speaking Proficiency Test measures some aspects of pronunciation and fluency, and the ability to recall basic Spanish words with some spontaneity.

Stanine norms are not available for Speaking. Broad student ratings, however, were determined in a few schools of the norms group under specially controlled conditions (1966). Corresponding to Ranges of Raw Scores, the suggested proficiency ratings in Speaking are shown in Table III.

TABLE III

SPEAKING RATINGS

Rating	Raw Score Range
Good	68-91
Fair	43-67
Poor	0-42

To facilitate interpretation in this study, mean scores for the parts of the Speaking Test and mean scores for the Total Speaking Test are shown in Table IV.

TABLE IV

SPEAKING

MEAN RAW SCORES

	X-Group (N-25)	C-Group (N-25)
Speaking (Total)	47.28	47.70
Vocabulary (Recall)	11.88	13.08
Pronunciation	25.28	28.04
Fluency	10.12*	6.58

Results

The t-tests for correlated means reveal a) no significant difference in the experimental and control groups in Total Speaking; b) the experimental group is significantly higher in Fluency. (*Level of significance: $p < .01$)[5]

Observations

Both mean scores in Total Speaking are close to the bottom of the fair range.

The selection of vocabulary in the experimental program is admittedly less extensive than in the control programs. Vocabulary is also more precisely controlled by considerations of immediate relevance to the students of Rincon High School.

Pronunciation per se may be a stronger objective in the control programs than in the experimental program. In the latter, pronunciation may be treated

---

[5] There is one chance in 100 of this being a chance difference. (See notes.)

more incidentally to Fluency.

Obviously the Fluency section of the Pimsleur Speaking Test does not precisely measure rapidity as a variable separate from the accuracy of response. (See recommendations.)

Final Conclusions

The study suggests that the experimental program may be providing ample opportunities for students to achieve the primary objective of the program - Fluency and Versatility without greatly penalizing them in the development of Listening and Reading objectives as compared with the other classes.

The study appears to confirm that the experimental students do less well in Writing than do the control students, probably as the result of a calculated lesser emphasis on Writing. The study also appears to confirm the wisdom of retaining students in an experimental program for longer than the first year, and may suggest that efforts should be made to strengthen their Writing proficiency before the experimental program is terminated.

Recommendations

The study should be continued through the second year using the experimental partners of the original matched-pairs who are still enrolled in the experimental program, and as many of the original control partners as continue in Spanish anywhere in District One. Losses of control partners may be made up from among others for whom pre-test data are available.

Mr. Valencia might find it advantageous to work with the Research Department in developing an instrument to measure his objective of Fluency and Versatility. The instrument would be useful to the district.

Group Test Results

The following is presented for the information of the participating teachers. Because of the uncontrolled variables, absolutely no valid conclusions may be drawn as far as comparative proficiency is concerned.[6]

TABLE V

GROUP TEST RESULTS

Teacher	1	2	3	4	5	Norms Group
N	47	40	23	36	41	5306
Pre-Test (Mean Stanines)						
Verbal Aptitude	4.83	5.77	5.00	3.92	4.71	5.00
Auditory Aptitude	6.34	6.90	7.00	5.50	6.73	5.00
Total Aptitude	5.72	6.65	6.17	4.86	5.95	5.00
Post-Test (Mean Stanines)						
Listening	5.11	4.78	5.57	4.64	6.12	5.00
Reading	4.91	5.18	6.04	4.25	6.56	5.00
Final Grade (Mean Grade)	2.67*	2.40	2.13	2.56	2.17	

*(N-46)

[6] I.e., differences may all be chance differences.

Correlation of Means, Separate Groups

Table VI shows the correlations between aptitudes, proficiencies, and final grades for the X-Group. Table VII shows the correlations for the C-Group.

The numbers from 1 to 8 running along the left-hand margin and along the top of the columns correspond to these factors:

1 - Verbal Language Aptitude

2 - Auditory Language Aptitude

3 - Total Language Aptitude

4 - Listening Proficiency

5 - Reading Proficiency

6 - Writing Proficiency

(The means of the above are expressed in Stanines.)

7 - Speaking Proficiency -- the mean is based upon the scale:

1 - good, 2 - fair, 3 - poor.

8 - Final Grade -- the mean is based upon the scale as used in Tucson Public Schools.

Tables VIII to XII show the correlations for the separate teacher groups, excluding from the N those students who are in the X- or C-Groups.

To find a correlation, e.g. for that between total language aptitude and final grade, read across from 3 (Total) to 8 (Final Grade) on Tables VI and VII, and from 3 (Total) to 6 (Final Grade) on Tables VIII - XII.

Coefficients of correlation ranging from zero to about .20 may be regarded as indicating no or negligible correlation.

Coefficients of correlation ranging from about .20 to .40 may be regarded as indicating a low degree of correlation.

Coefficients of correlation ranging from about .40 to .60 may be regarded as indicating a moderate degree of correlation.

Coefficients of correlation ranging from about .60 to .80 may be regarded as indicating a marked degree of correlation.

Coefficients of correlation ranging from about .80 to 1.00 may be regarded as indicating high correlation. (From A Primer of Statistics, Franzblau, p. 81.)

As can be seen from the tables, there are many significant positive relationships between the factors.

It must be emphasized, however, that no conclusions are to be drawn from this information.

CORRELATION
1 VER 2 AUD 3 TOT 4 LSTN 5 RD 6 WRT 7 SPK 8 FINAL GR

APTITUDE, ACHIEVEMENT, AND FINAL GRADE

23 AUGUST 1968 X - GROUP

	1	2	3	4	5	6	7	8
N	25.	25.	25.	25.	25.	25.	25.	25.
MEAN	5.32	6.80	6.44	4.40	4.80	4.24	2.28	2.52
S.D.	1.52	1.47	1.19	1.44	1.53	1.13	0.46	1.08
1	0.00	0.35	0.70	0.49	0.35	0.71	±0.49	±0.66
2	4.33	0.00	0.74	0.33	0.37	0.46	±0.41	±0.53
3	5.13	1.81	0.00	0.47	0.42	0.72	±0.62	±0.73
4	3.07	7.13	7.45	0.00	0.25	0.50	±0.24	±0.32
5	1.50	5.94	5.48	1.10	0.00	0.44	±0.21	±0.66
6	5.01	9.23	12.70	0.61	1.94	0.00	±0.54	±0.72
7	8.49	13.21	13.70	6.56	7.47	6.86	0.00	±0.62
8	5.88	9.53	9.25	4.55	4.77	4.19	1.36	0.00
0	0.30	0.29	0.24	0.29	0.31	0.23	0.09	0.22

TABLE VI

CORRELATION
1 VER 2 AUD 3 TOT 4 LSTN 5 RD 6 WRT 7 SPK 8 FINAL GR
APTITUDE, ACHIEVEMENT, AND FINAL GRADE
23 AUGUST 1968 C-GROUP

	1	2	3	4	5	6	7	8
N	25.	25.	25.	25.	25.	25.	25.	25.
MEAN	5.20	6.76	6.28	5.28	4.80	5.24	2.28	2.16
S.D.	1.44	1.36	1.17	1.81	2.02	1.48	0.61	0.90
1	0.00	0.03	0.75	0.22	0.16	0.46	+0.44	+0.41
2	3.98	0.00	0.54	-0.07	-0.02	-0.11	+0.12	-0.34
3	5.66	1.95	0.00	0.18	0.29	0.30	+0.46	+0.28
4	0.19	3.15	2.53	0.00	0.62	0.67	+0.48	+0.62
5	0.87	3.99	3.66	1.42	0.00	0.62	+0.46	+0.60
6	0.13	3.58	3.26	0.15	1.37	0.00	+0.67	+0.69
7	8.11	14.38	12.87	6.88	5.33	7.60	0.00	+0.52
8	7.64	16.99	12.36	6.31	4.96	7.01	0.77	0.00
0	0.29	0.27	0.23	0.36	0.40	0.30	0.12	0.18

CORRELATION
1 VERB 2 AUD 3 TOT 4 LISTEN 5 READ 6 FINAL GR
APTITUDE, ACHIEVEMENT, AND FINAL GRADE
23 AUGUST 1968 TEACHER 1

	1	2	3	4	5	6
N	37.	37.	37.	37.	37.	37.
MEAN	4.59	6.22	5.51	5.11	5.11	2.78
S.D.	1.39	1.93	1.55	2.01	1.84	1.18
1	0.00	0.20	0.79	0.23	0.40	+0.49
2	4.08	0.00	0.63	0.21	0.13	+0.22
3	4.80	2.77	0.00	0.19	0.33	+0.56
4	1.29	2.72	1.07	0.00	0.52	+0.39
5	1.53	2.70	1.25	0.00	0.00	+0.55
6	4.12	3.43	5.85	5.24	5.28	0.00
D	0.31	0.32	0.25	0.33	0.30	0.19

TABLE VIII

CORRELATION
1 VERB 2 AUD 3 TOT 4 LISTEN 5 READ 6 FINAL GR
APTITUDE, ACHIEVEMENT, AND FINAL GRADE
23 AUGUST 1968 TEACHER 2

	1	2	3	4	5	6
N	15.	15.	15.	15.	15.	15.
MEAN	6.53	7.07	7.00	5.40	5.80	2.20
S.D.	2.10	1.67	1.73	2.03	2.18	1.21
1	0.00	0.72	0.90	0.67	0.84	+0.72
2	1.42	0.00	0.87	0.82	0.75	+0.72
3	1.97	0.29	0.00	0.81	0.87	+0.82
4	2.61	5.49	5.24	0.00	0.65	+0.76
5	2.32	3.41	4.29	0.88	0.00	+0.58
6	5.44	7.06	6.62	4.06	4.58	0.00
0	0.54	0.43	0.45	0.52	0.56	0.31

TABLE IX

CORRELATIONS
1 VERB 2 AUD 3 TOT 4 LISTEN 5 READ 6 FINAL GR
APTITUDE, ACHIEVEMENT, AND FINAL GRADE
23 AUGUST 1968 TEACHER 3

	1	2	3	4	5	6
N	19.	19.	19.	19.	19.	19.
MEAN	4.84	7.05	6.05	5.53	5.95	2.26
S.D.	1.74	1.84	1.78	1.35	1.81	1.05
1	0.00	0.59	0.88	0.13	0.33	+0.62
2	5.95	0.00	0.85	0.19	0.18	+0.50
3	6.17	4.36	0.00	0.27	0.38	+0.72
4	1.45	3.22	1.19	0.00	0.54	+0.34
5	2.35	2.07	0.23	1.17	0.00	+0.49
6	4.45	8.25	6.26	7.23	6.43	0.00
0	0.40	0.42	0.41	0.31	0.42	0.24

TABLE X

CORRELATION
1 VERB 2 AUD 3 TOT 4 LISTEN 5 READ 6 FINAL GR .
APTITUDE, ACHIEVEMENT, AND FINAL GRADE
23 AUGUST 1968 TEACHER 4

	1	2	3	4	5	6
N	31.	31.	31.	31.	31.	31.
MEAN	3.90	5.35	4.77	4.81	4.45	2.61
S.D.	1.80	1.78	1.45	1.64	2.05	0.99
1	0.00	0.19	0.72	0.08	0.11	+0.30
2	3.55	0.00	0.69	0.23	0.08	+0.37
3	3.85	2.47	0.00	0.19	0.13	+0.53
4	2.16	1.44	0.09	0.00	0.38	+0.48
5	1.19	1.93	0.76	0.95	0.00	+0.40
6	3.13	6.53	5.61	5.34	3.92	0.00
0	0.32	0.32	0.26	0.29	0.37	0.18

TABLE XI

CORRELATION
1 VERB 2 AUD 3 TOT 4 LISTEN 5 READ 6 FINAL GR
APTITUDE, ACHIEVEMENT, AND FINAL GRADE
23 AUGUST 1968 TEACHER 5

	1	2	3	4	5	6
N	34.	34.	34.	34.	34.	34.
MEAN	4.68	6.71	5.91	6.03	6.56	2.26
S.D.	1.87	1.61	1.60	2.04	1.91	1.19
1	0.00	0.53	0.80	0.23	0.35	+0.42
2	6.98	0.00	0.80	0.48	0.41	+0.39
3	6.38	4.59	0.00	0.35	0.37	+0.50
4	3.24	2.07	0.33	0.00	0.65	+0.52
5	5.09	0.45	1.90	1.86	0.00	+0.56
6	5.39	11.08	8.78	7.73	9.08	0.00
0	0.32	0.28	0.27	0.35	0.33	0.20

TABLE XII

Notes

Statistical Significance

The report takes care to emphasize that some of the findings of this experiment are statistically significant, while others are not. Since some of the statistically insignificant differences may appear rather great, a further word concerning statistical significance may be in order. The following notes are taken from A Psycholinguistic Experiment in Foreign-Language Teaching, by George A. C. Scherer and Michael Wertheimer:

> The "Statistical significance" of a statistic is inversely related to the likelihood that the statistic reflects a chance event; the greater the statistical significance, the less likely it is that the correlation, difference, etc., obtained in a series of measurements is untrustworthy or unreliable. The greater the correlation or the larger the difference, etc., other things being equal, the more likely it is that the correlation or difference is repeatable, "real," or "statistically significant."
>
> The .05 level, .01 level, etc., refer to the statistical significance of a finding. If there is one chance in a hundred that the finding is due to chance, then the finding is called significant at the .01 level, or "very significant"; one chance in twenty defines the .05 level. If there is one chance in five, for instance, that the finding is due to chance, the finding is statistically not significant. Perhaps an example may...help to clarify. Assume that the true correlation...between mathematical ability and blackness of hair is zero. Now if we measure mathematical ability and blackness of hair on a sample of 20 people, we might obtain a correlation of .08 in our sample. Another sample might yield, say, -.03, still another .12, and so on. These are presumably chance variations around a correlation of 0.00, the "true" one in the population of all people. But since a correlation of .68 or -.73 would be extremely unlikely to occur in a sample if the true r is zero, we should hesitate to conclude with such a large r[7] that the true r is zero--in other words, an r of .68 or -.73 would be likely to be statistically significant, i.e. unlikely to occur by chance. In all

---

[7] r = coefficient of correlation

our treatment of correlation coefficients here, statistical significance refers to the likelihood of obtaining a sample r of this size or larger if there really is zero correlation in the population. Thus in general the greater the obtained correlation coefficient, the more likely it is to be statistically significant." (pp. 70-71)

Standard Deviation (S D)

The standard deviation (σ) of a set of scores provides an index of the variability, scatter, or dispersion of those scores. Mathematically, it is the square root of the mean squared deviation of the scores from their own mean. The larger the σ , the greater the variability, that is, the more dispersed the scores; the smaller the σ , the less variability, that is, the more "compact" the distribution of scores. In addition to providing an efficient description of variability, the standard deviation has some very useful statistical properties, and enters, among other things, into the computation of the r, the Pearson product-moment correlation coefficient, and into various significance tests, such as t.

The t Test

The t test is a standard statistical device for assessing the significance...of a finding. In one form, it can be considered as an index of the size of an obtained difference relative to the variability of measurement, or of "signal" relative to "noise." The larger the t obtained in testing a given difference, the less likely it is that the difference is due to chance alone, i.e., the more likely it is that the difference will be statistically significant; the smaller the t the more likely it is that the obtained difference is just a chance fluctuation, and that there is really no difference in the population.

The statistical significance of the difference between the means of the groups, as indicated by the t test, is tabled as p, the probability that the difference is due to chance. Statistical convention generally considers any p of .06 or greater as indicating that the difference is not significant, not trustworthy, not established; p of .05 or lower is generally considered significant, .01 or lower, very significant. (p. 71)