Paper presented at LTRC 2015 annual conference

Analysis of Norm-Referencing Modifiers as a Component of Rating Rubrics

Alex Volkov, Jake Stone, Taylor Asbury, Michael Hsu

Paragon Testing Enterprises

University of British Columbia

Agenda

- ✓ Rationale
- ✓ Research questions
- ✓ Methodology
- ✓ Analysis
- ✓ Conclusion

Rating Scale Design

✓ Empirically informed rating rubric

✓ How do modifiers function as a component of rating rubrics?

✓ Lack of research in modifiers

Research Questions

✓ Do modifiers correspond to a specific ability range?

✓ Will the same modifier attached to different descriptors nonetheless be targeting a similar ability range?

Descriptors and Modifiers

- ✓ Creates a cohesive text
- ✓ Grammar is correct
- ✓ Punctuation is correct
- ✓ Vocabulary is appropriate
- ✓ Writing is intelligible

- ✓ Partially
- ✓ Sufficiently
- ✓ Mostly

Creates a partially cohesive text

Creates a sufficiently cohesive text

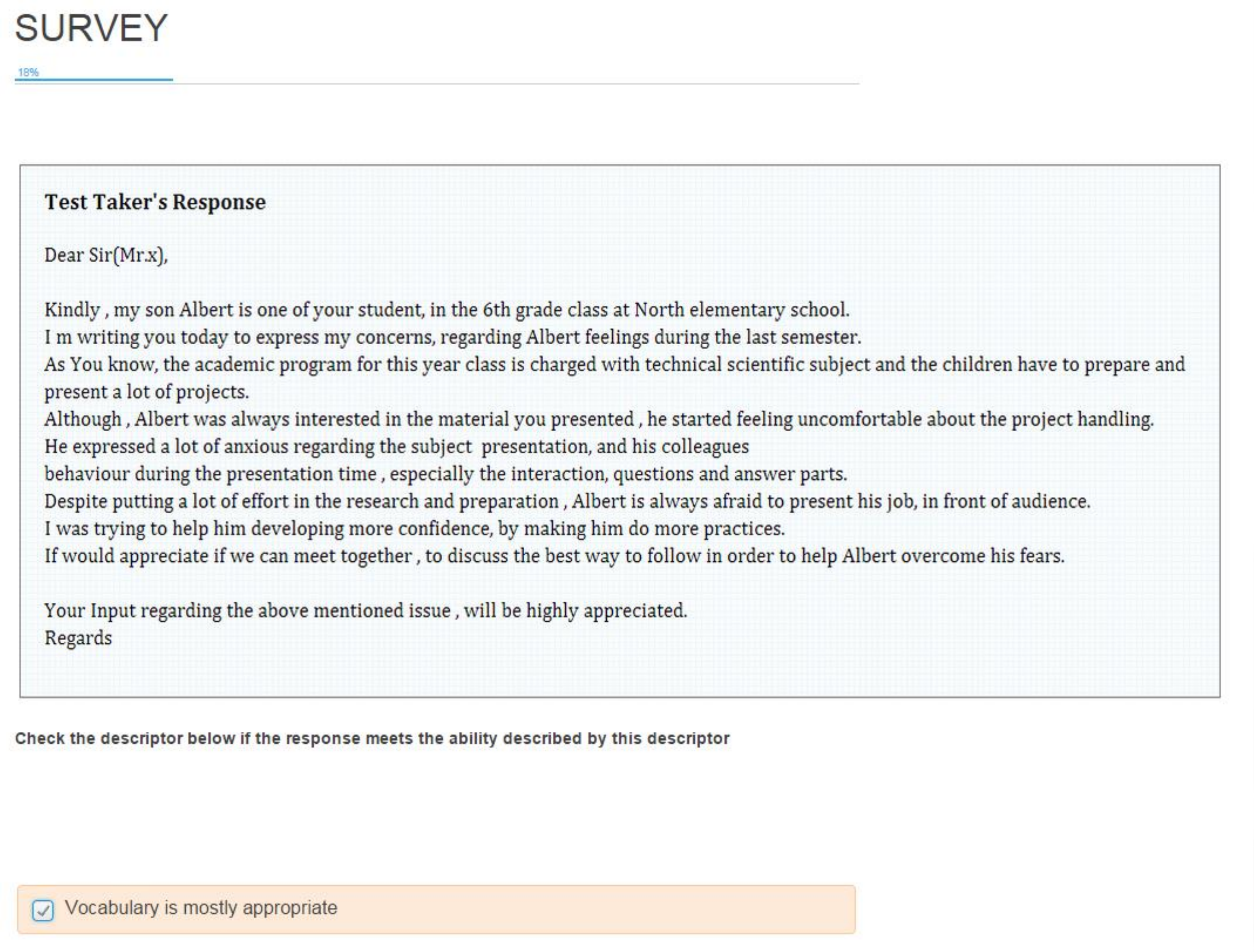
Creates a mostly cohesive text

Methodology

✓ 30 Samples, 10 experienced raters, 5 descriptor types, 3 modifiers. Fully crossed design

✓ Online rating using Fluid Surveys

Fluid Surveys



SURVEY

18%

Test Taker's Response

Dear Sir(Mr.x),

Kindly , my son Albert is one of your student, in the 6th grade class at North elementary school.
I m writing you today to express my concerns, regarding Albert feelings during the last semester.
As You know, the academic program for this year class is charged with technical scientific subject and the children have to prepare and present a lot of projects.
Although , Albert was always interested in the material you presented , he started feeling uncomfortable about the project handling.
He expressed a lot of anxious regarding the subject presentation, and his colleagues
behaviour during the presentation time , especially the interaction, questions and answer parts.
Despite putting a lot of effort in the research and preparation , Albert is always afraid to present his job, in front of audience.
I was trying to help him developing more confidence, by making him do more practices.
If would appreciate if we can meet together , to discuss the best way to follow in order to help Albert overcome his fears.

Your Input regarding the above mentioned issue , will be highly appreciated.
Regards

Check the descriptor below if the response meets the ability described by this descriptor

☑ Vocabulary is mostly appropriate

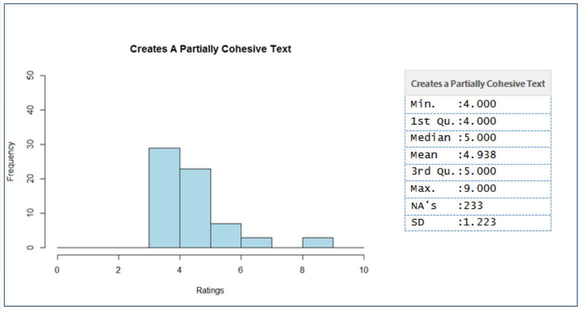
Analysis

✓ Descriptive statistics

✓ ANOVA

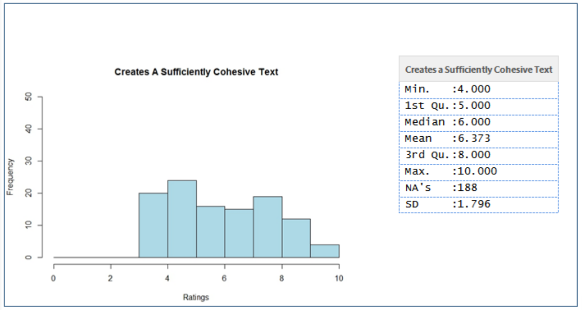
✓ MFRM

Analysis: Descriptive Statistics



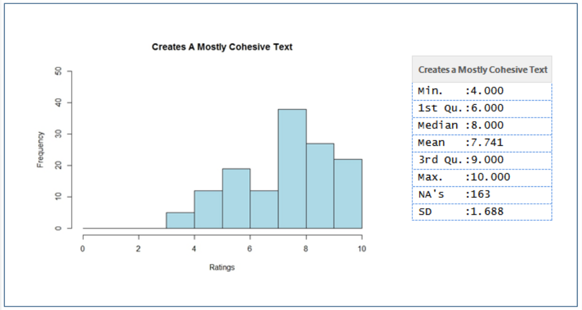
```
Creates a Partially Cohesive Text
Min.   :4.000
1st Qu.:4.000
Median :5.000
Mean   :4.938
3rd Qu.:5.000
Max.   :9.000
NA's   :233
SD     :1.223
```

✓ Creates a **partially** cohesive text



```
Creates a Sufficiently Cohesive Text
Min.   :4.000
1st Qu.:5.000
Median :6.000
Mean   :6.373
3rd Qu.:8.000
Max.   :10.000
NA's   :188
SD     :1.796
```

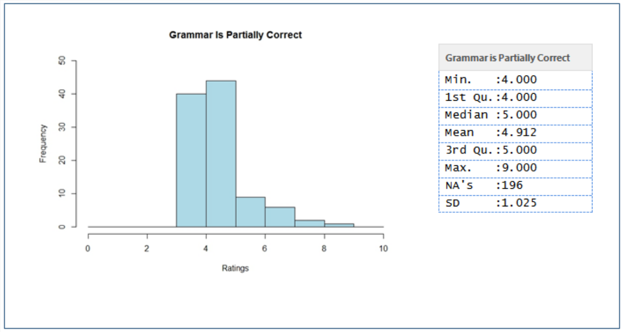
✓ Creates a **sufficiently** cohesive text



```
Creates a Mostly Cohesive Text
Min.   :4.000
1st Qu.:6.000
Median :8.000
Mean   :7.741
3rd Qu.:9.000
Max.   :10.000
NA's   :163
SD     :1.688
```

✓ Creates a **mostly** cohesive text

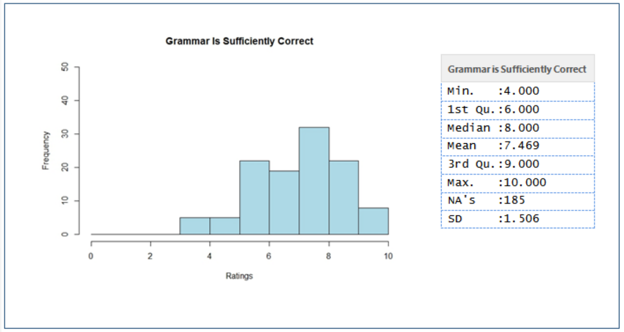
Analysis: Descriptive Statistics



Grammar is Partially Correct

```
Min.   :4.000
1st Qu.:4.000
Median :5.000
Mean   :4.912
3rd Qu.:5.000
Max.   :9.000
NA's   :196
SD     :1.025
```

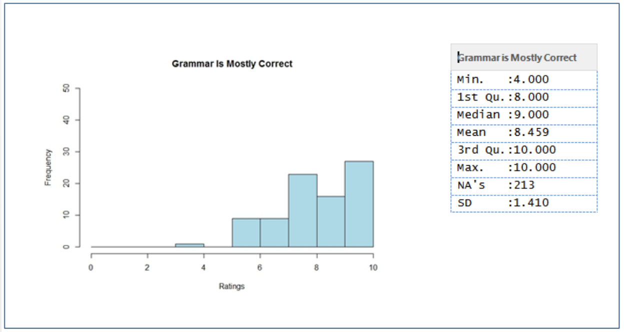
✓ Grammar is **partially** correct



Grammar is Sufficiently Correct

```
Min.   :4.000
1st Qu.:6.000
Median :8.000
Mean   :7.469
3rd Qu.:9.000
Max.   :10.000
NA's   :185
SD     :1.506
```

✓ Grammar is **sufficiently** correct



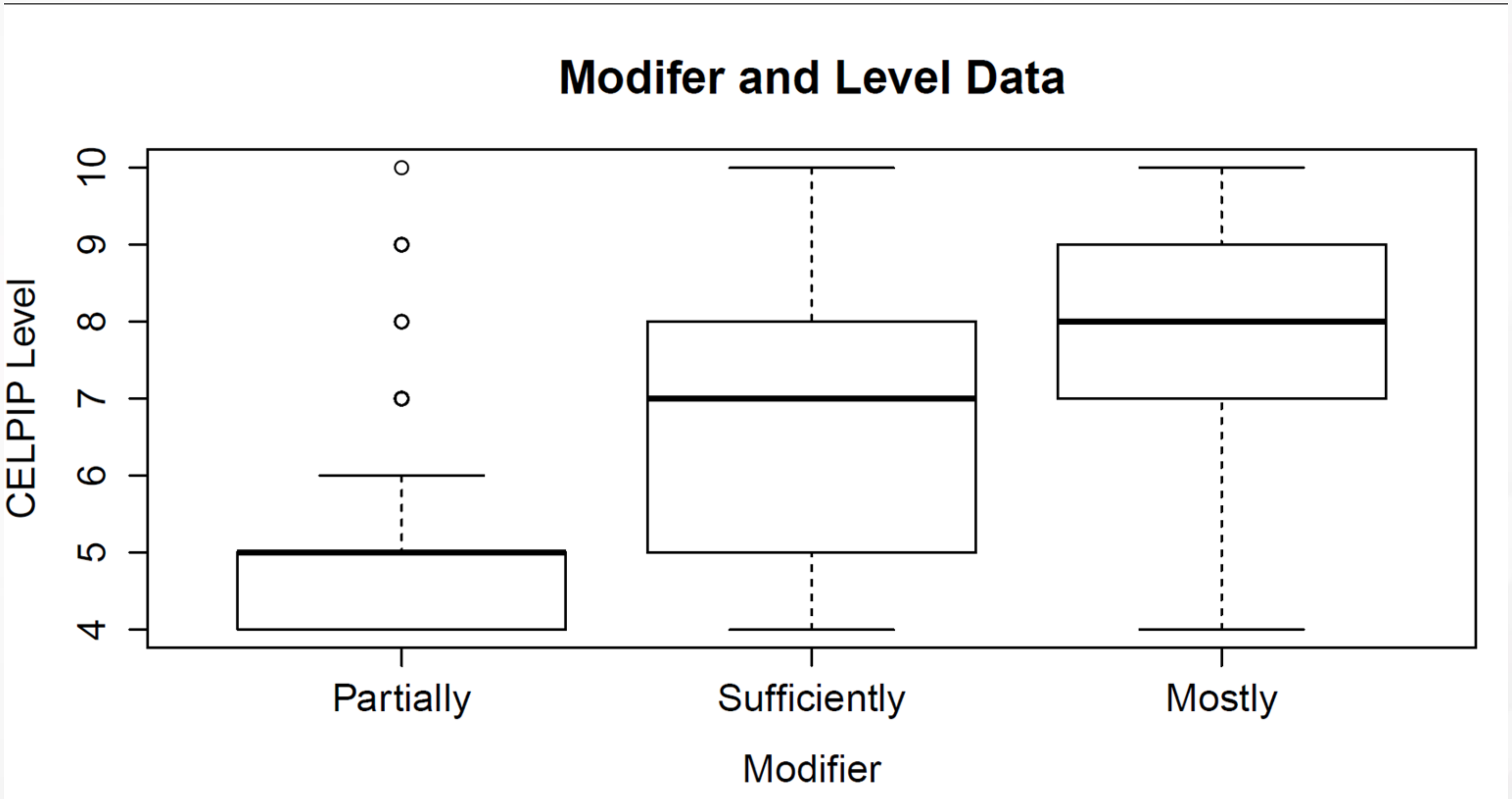
Grammar is Mostly Correct

```
Min.   :4.000
1st Qu.:8.000
Median :9.000
Mean   :8.459
3rd Qu.:10.000
Max.   :10.000
NA's   :213
SD     :1.410
```

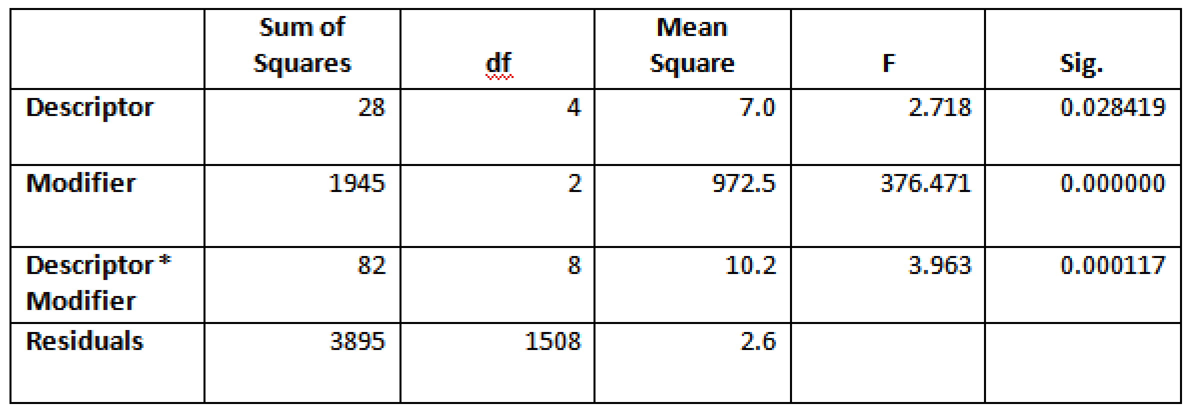
✓ Grammar is **mostly** correct

TESTING ENTERPRISES
Paragon

Analysis: Box Plots
(across all descriptors)



Analysis: ANOVA



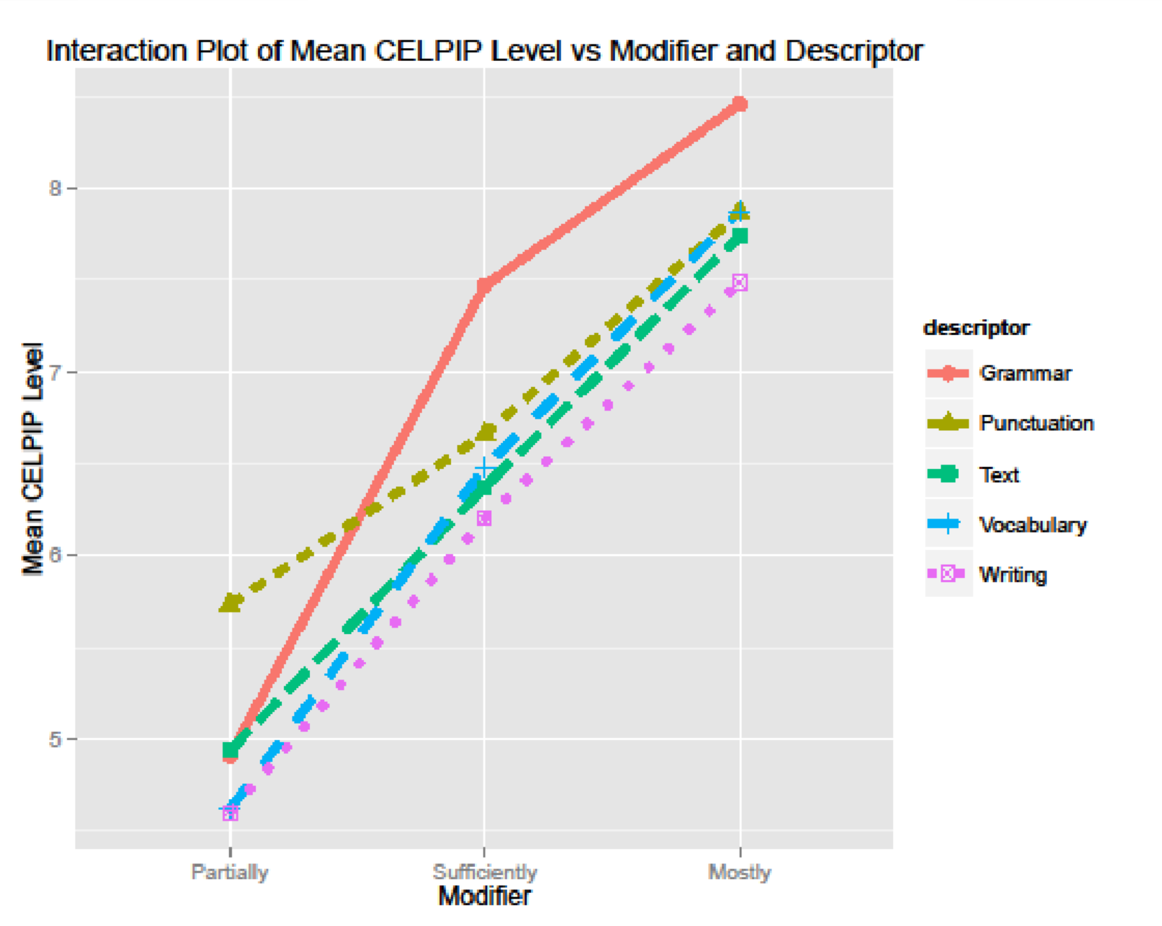
	Sum of Squares	df	Mean Square	F	Sig.
Descriptor	28	4	7.0	2.718	0.028419
Modifier	1945	2	972.5	376.471	0.000000
Descriptor * Modifier	82	8	10.2	3.963	0.000117
Residuals	3895	1508	2.6		

✓ Descriptor type

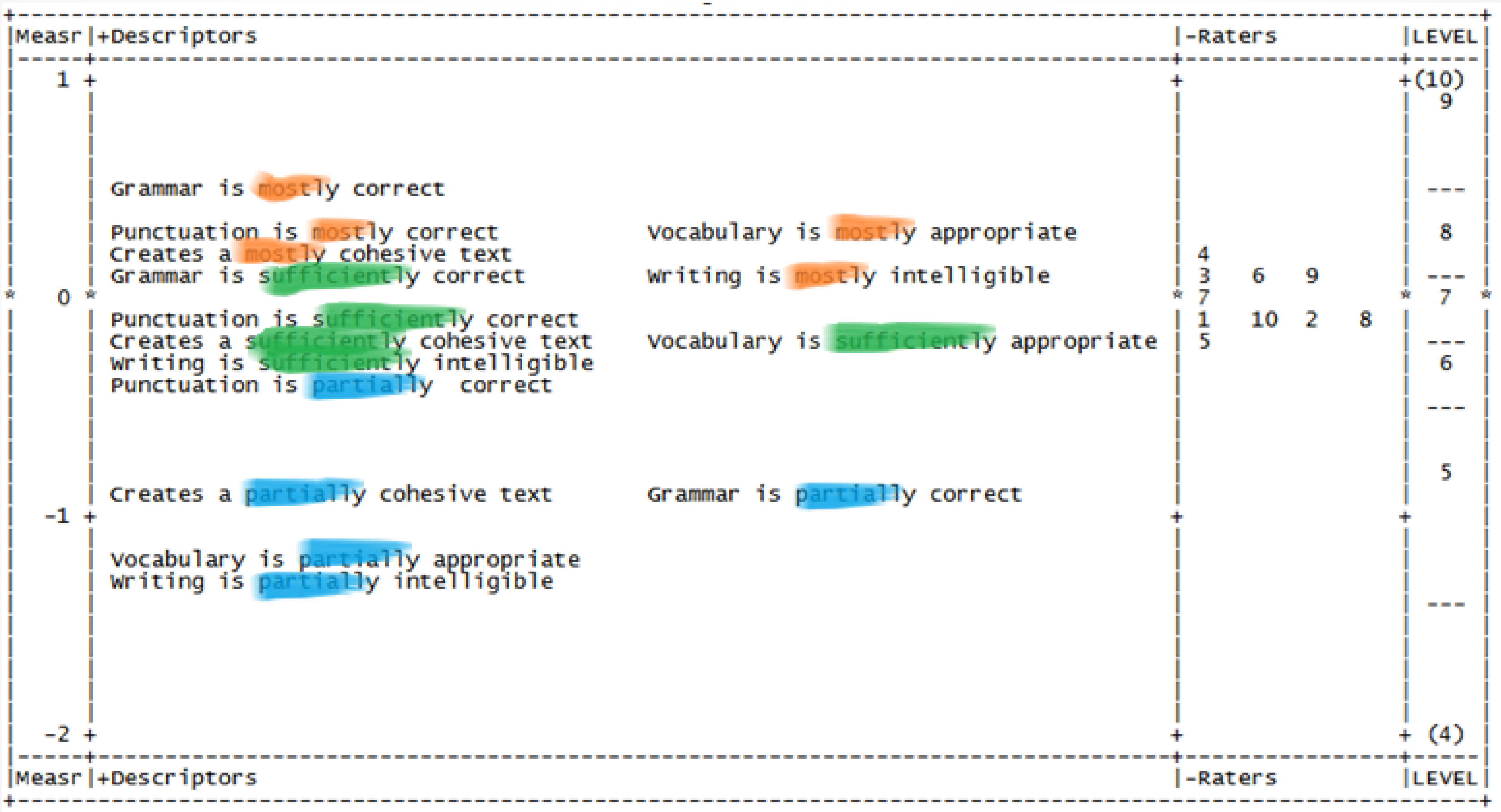
✓ Modifier

✓ Statistical significance

Analysis: ANOVA



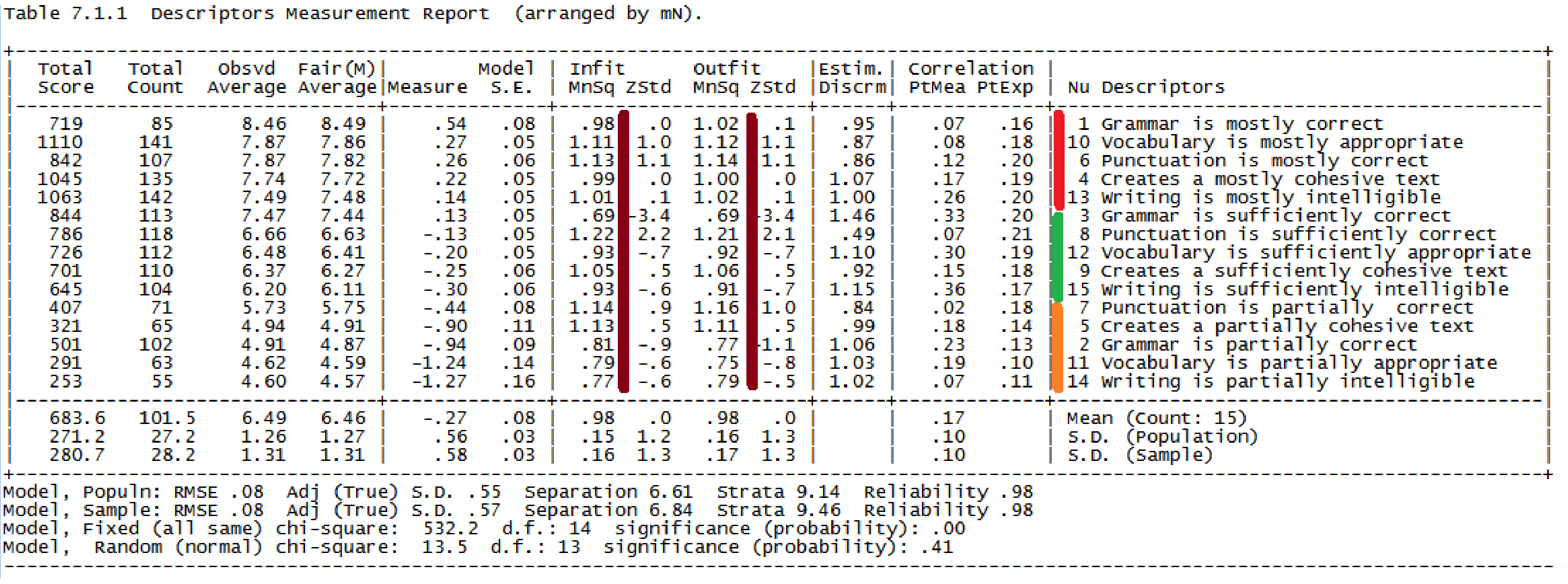
Analysis: MFRM



```
+---------------------------------------------------------------------------------------------------------------+
|Measr|+Descriptors                                                                     |-Raters          |LEVEL|
|-----+---------------------------------------------------------------------------------+-----------------+-----|
|   1 +                                                                                 +                 +(10) |
|     |                                                                                 |                 |  9  |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |     |
|     | Grammar is mostly correct                                                       |                 | --- |
|     |                                                                                 |                 |     |
|     | Punctuation is mostly correct           Vocabulary is mostly appropriate        |                 |  8  |
|     | Creates a mostly cohesive text                                                  | 4               |     |
|     | Grammar is sufficiently correct         Writing is mostly intelligible          | 3   6   9       | --- |
*   0 *                                                                                 * 7               *  7  *
|     | Punctuation is sufficiently correct                                             | 1   10  2   8   |     |
|     | Creates a sufficiently cohesive text    Vocabulary is sufficiently appropriate  | 5               | --- |
|     | Writing is sufficiently intelligible                                            |                 |  6  |
|     | Punctuation is partially  correct                                               |                 |     |
|     |                                                                                 |                 | --- |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |  5  |
|     | Creates a partially cohesive text       Grammar is partially correct            |                 |     |
|  -1 +                                                                                 +                 +     |
|     |                                                                                 |                 |     |
|     | Vocabulary is partially appropriate                                             |                 |     |
|     | Writing is partially intelligible                                               |                 |     |
|     |                                                                                 |                 | --- |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |     |
|     |                                                                                 |                 |     |
|  -2 +                                                                                 +                 + (4) |
|-----+---------------------------------------------------------------------------------+-----------------+-----|
|Measr|+Descriptors                                                                     |-Raters          |LEVEL|
+---------------------------------------------------------------------------------------------------------------+
```

TESTING ENTERPRISES Paragon

Analysis. MFRM



```
Table 7.1.1  Descriptors Measurement Report  (arranged by mN).

+-----------------------------------------------------------------------------------------------------------------------------------------+
|  Total    Total    Obsvd  Fair(M)|        Model | Infit       Outfit     |Estim.| Correlation |                                         |
|  Score    Count  Average  Average|Measure  S.E. | MnSq ZStd   MnSq ZStd  |Discrm| PtMea PtExp |  Nu Descriptors                         |
|----------------------------------+--------------+------------------------+------+-------------+-----------------------------------------|
|   719      85      8.46    8.49  |    .54    .08 |   .98   .0   1.02   .1 |  .95 |   .07   .16 |   1 Grammar is mostly correct           |
|  1110     141      7.87    7.86  |    .27    .05 |  1.11  1.0   1.12  1.1 |  .87 |   .08   .18 |  10 Vocabulary is mostly appropriate    |
|   842     107      7.87    7.82  |    .26    .06 |  1.13  1.1   1.14  1.1 |  .86 |   .12   .20 |   6 Punctuation is mostly correct       |
|  1045     135      7.74    7.72  |    .22    .05 |   .99   .0   1.00   .0 | 1.07 |   .17   .19 |   4 Creates a mostly cohesive text      |
|  1063     142      7.49    7.48  |    .14    .05 |  1.01   .1   1.02   .1 | 1.00 |   .26   .20 |  13 Writing is mostly intelligible      |
|   844     113      7.47    7.44  |    .13    .05 |   .69 -3.4    .69 -3.4 | 1.46 |   .33   .20 |   3 Grammar is sufficiently correct     |
|   786     118      6.66    6.63  |   -.13    .05 |  1.22  2.2   1.21  2.1 |  .49 |   .07   .21 |   8 Punctuation is sufficiently correct |
|   726     112      6.48    6.41  |   -.20    .05 |   .93  -.7    .92  -.7 | 1.10 |   .30   .19 |  12 Vocabulary is sufficiently appropriate |
|   701     110      6.37    6.27  |   -.25    .06 |  1.05   .5   1.06   .5 |  .92 |   .15   .18 |   9 Creates a sufficiently cohesive text |
|   645     104      6.20    6.11  |   -.30    .06 |   .93  -.6    .91  -.7 | 1.15 |   .36   .17 |  15 Writing is sufficiently intelligible |
|   407      71      5.73    5.75  |   -.44    .08 |  1.14   .9   1.16  1.0 |  .84 |   .02   .18 |   7 Punctuation is partially  correct   |
|   321      65      4.94    4.91  |   -.90    .11 |  1.13   .5   1.11   .5 |  .99 |   .18   .14 |   5 Creates a partially cohesive text   |
|   501     102      4.91    4.87  |   -.94    .09 |   .81  -.9    .77 -1.1 | 1.06 |   .23   .13 |   2 Grammar is partially correct        |
|   291      63      4.62    4.59  |  -1.24    .14 |   .79  -.6    .75  -.8 | 1.03 |   .19   .10 |  11 Vocabulary is partially appropriate |
|   253      55      4.60    4.57  |  -1.27    .16 |   .77  -.6    .79  -.5 | 1.02 |   .07   .11 |  14 Writing is partially intelligible   |
|----------------------------------+--------------+------------------------+------+-------------+-----------------------------------------|
|   683.6   101.5    6.49    6.46  |   -.27    .08 |   .98   .0    .98   .0 |      |   .17       | Mean (Count: 15)                        |
|   271.2    27.2    1.26    1.27  |    .56    .03 |   .15  1.2    .16  1.3 |      |   .10       | S.D. (Population)                       |
|   280.7    28.2    1.31    1.31  |    .58    .03 |   .16  1.3    .17  1.3 |      |   .10       | S.D. (Sample)                           |
+-----------------------------------------------------------------------------------------------------------------------------------------+
Model, Populn: RMSE .08  Adj (True) S.D. .55  Separation 6.61  Strata 9.14  Reliability .98
Model, Sample: RMSE .08  Adj (True) S.D. .57  Separation 6.84  Strata 9.46  Reliability .98
Model, Fixed (all same) chi-square:  532.2  d.f.: 14  significance (probability): .00
Model,  Random (normal) chi-square:  13.5  d.f.: 13  significance (probability): .41
-------------------------------------------------------------------------------------------------------------------------------------------
```

Feedback from Raters

- ✓ Even though the raters were not told the purpose of the study, they noticed the systematic use of the modifiers
- ✓ Many raters informed us that limiting the number of modifiers to 3 was very helpful
- ✓ Each rater began to devise heuristics for judging each modifier

Future work

- ✓ Can training or seminars help raters develop a shared consensus as to how different performance levels correspond with different modifiers?
- ✓ Can we discern a limited and highly descriptive pool of modifiers for systematic application?

Conclusion

✓ Modifiers can be discerning and can have a major effect on raters’ perceptions

✓ Modifiers should be carefully selected. Some are better targeted at a specific ability range.

✓ Using a limited number of modifiers systematically may help with inter-rater reliability

Thank you!

Analysis of Norm-Referencing Modifiers as a Component of Rating Rubrics

Questions?