Investments in Accounting Resources and the Implications for External Reporting and Disclosure

Internet Appendix

December 2019

Table of Contents

IA.1 Additional Descriptive Statistics: Grammatical Violations

We provide descriptive statistics of our measure (GV) by year, by type, and by industry. Table IA.1.1 details the average of GV by year. The average number of grammatical violations is stable across the sample period. Table IA.1.2 describes the average of GV split into the six categories of grammatical rules as classified by the *LanguageTool* project. The rules are commonly confused words (e.g., affect vs effect; dependent vs. dependant, etc.), grammatical errors (e.g., adverb put before the verb 'are'; adverb 'often' put after vs. put before the verb 'are', etc.), miscellaneous ('an' instead of 'a'), punctuation (e.g., comma after a conjunctive/linking adverb; space before a full stop), and redundant phrases affecting the conciseness of the text (e.g., 'past experience'; 'period of time', etc.).[1] In Table IA.1.3, we show the distribution of grammatical violations by type and industry. Although the distribution of error types varies by year and industry, the distribution is fairly consistent across the sample.

[1] We also detect that less than 4% of our sample contains errors such as collocation, non-standard phrases, and semantics that we suppress from our descriptive statistics due to the small number of observations.

Table IA.1.1: Distribution of Grammatical Violations (GVs) by Year

Year	N. Obs	GV	% Change
1993	57	0.230	
1994	474	0.204	11%
1995	1,440	0.203	0%
1996	2,604	0.197	3%
1997	2,781	0.210	-7%
1998	2,771	0.221	-5%
1999	2,813	0.222	0%
2000	3,014	0.189	15%
2001	2,964	0.207	-10%
2002	3,959	0.209	-1%
2003	3,959	0.227	-9%
2004	3,862	0.236	-4%
2005	4,187	0.204	14%
2006	4,216	0.203	1%
2007	4,199	0.200	2%
2008	4,399	0.218	-9%
2009	4,255	0.219	0%
2010	4,315	0.212	3%
2011	4,518	0.231	-9%
2012	4,431	0.206	11%
2013	4,413	0.216	-4%
2014	4,458	0.215	0%
2015	4,232	0.187	13%
2016	4,237	0.182	3%
2017	4,134	0.189	-4%
2018	3,885	0.183	3%

Table IA.1.2: Distribution of Grammatical Violations by Type

Year	Commonly Confused Words	% Change	Grammar	% Change	Miscellaneous	% Change	Punctuation Errors	% Change	Redundant Phrases	% Change
1993	0.0185		0.0018		0.0181		0.1280		0.0380	
1994	0.0055	70%	0.0004	76%	0.0208	-15%	0.1039	19%	0.0492	-29%
1995	0.0076	-39%	0.0006	-38%	0.0208	0%	0.0970	7%	0.0529	-8%
1996	0.0109	-43%	0.0005	22%	0.0193	7%	0.0823	15%	0.0623	-18%
1997	0.0216	-99%	0.0008	-63%	0.0192	1%	0.0755	8%	0.0708	-14%
1998	0.0266	-23%	0.0013	-63%	0.0175	9%	0.0748	1%	0.0826	-17%
1999	0.0188	29%	0.0015	-20%	0.0184	-5%	0.0819	-9%	0.0817	1%
2000	0.0135	28%	0.0015	5%	0.0182	1%	0.0751	8%	0.0681	17%
2001	0.0161	-19%	0.0015	-2%	0.0186	-2%	0.0815	-9%	0.0797	-17%
2002	0.0152	6%	0.0018	-21%	0.0159	15%	0.0840	-3%	0.0841	-6%
2003	0.0144	6%	0.0044	-147%	0.0147	8%	0.0869	-3%	0.0973	-16%
2004	0.0127	12%	0.0038	14%	0.0225	-54%	0.0908	-4%	0.1057	-9%
2005	0.0103	19%	0.0018	53%	0.0165	26%	0.0789	13%	0.0910	14%
2006	0.0094	9%	0.0019	-7%	0.0170	-3%	0.0768	3%	0.0928	-2%
2007	0.0084	11%	0.0018	8%	0.0174	-2%	0.0746	3%	0.0925	0%
2008	0.0103	-23%	0.0016	11%	0.0167	4%	0.0784	-5%	0.1030	-11%
2009	0.0098	5%	0.0018	-17%	0.0191	-14%	0.0788	0%	0.1008	2%
2010	0.0074	24%	0.0020	-10%	0.0187	2%	0.0763	3%	0.0996	1%
2011	0.0063	15%	0.0022	-10%	0.0174	7%	0.0810	-6%	0.1151	-16%
2012	0.0049	23%	0.0024	-10%	0.0194	-11%	0.0719	11%	0.0983	15%
2013	0.0032	34%	0.0026	-10%	0.0206	-6%	0.0710	1%	0.1088	-11%
2014	0.0024	26%	0.0046	-75%	0.0182	12%	0.0682	4%	0.1112	-2%
2015	0.0022	9%	0.0026	44%	0.0181	0%	0.0616	10%	0.0966	13%
2016	0.0019	12%	0.0026	2%	0.0187	-3%	0.0565	8%	0.0948	2%
2017	0.0015	23%	0.0026	0%	0.0230	-23%	0.0608	-8%	0.1005	-6%
2018	0.0013	12%	0.0026	-1%	0.0198	14%	0.0491	19%	0.1015	-1%

Table IA.1.3: Distribution of Grammatical Violations by Industry

This table presents the breakdown of GV by type (Grammar, Inconsistency, Misspelling, Style and Typographical). Industry 6 (Business Equipment) comprises of Computers, Software, and Electronic Equipment and industry 12 (Others) is comprised of Mines, Construction, Building Materials, Transportation, Hotels, Business Services, Entertainment.

Ind.	N. Obs	Industry	Commonly Confused Words	Grammar	Miscellaneous	Punctuation Errors	Redundant Phrases
1	2,810	Non-Durables	0.0115	0.0026	0.0191	0.0756	0.0633
2	670	Consumer Durables	0.0099	0.0008	0.0119	0.0759	0.1003
3	7,151	Manufacturing	0.0098	0.0025	0.0182	0.0757	0.0842
4	4,334	Energy	0.0085	0.0028	0.0305	0.0758	0.0899
5	2,034	Chemicals	0.0085	0.0021	0.0180	0.0765	0.0998
6	12,316	Business Equipment	0.0129	0.0023	0.0171	0.0741	0.1030
7	2,009	Telecommunication	0.0089	0.0021	0.0141	0.0843	0.0633
8	3,946	Utilities	0.0044	0.0011	0.0195	0.0736	0.0758
9	5,082	Shops	0.0099	0.0021	0.0192	0.0769	0.0885
10	12,524	Healthcare	0.0078	0.0019	0.0163	0.0648	0.0983
11	18,443	Finance	0.0077	0.0023	0.0175	0.0728	0.1046
12	19,350	Other	0.0092	0.0024	0.0195	0.0816	0.0919

IA.2 Discretionary Accruals and Grammatical Violations

We document that grammatical violations are correlated with FRQ using future restatements and internal control weaknesses (Tables 4 and 5, respectively). We report our results using an alternative measure of FRQ, discretionary accruals, in Table IA.2. In columns (1) and (2), we regress GV on two measures of discretionary accruals: the absolute value of discretionary accruals using the Jones model and the modified Jones model, respectively. Across both columns, the coefficient on GV is positive and significant. Thus, grammatical errors are positively related to low financial reporting quality, as measured by discretionary accrual models.

Table IA.2: Discretionary Accruals and Grammatical Violations

The dependent variable is ABS JONES and ABS MOD JONES that measures the discretionary accruals following the Jones model in Model (1) and the Modified Jones model in Model (2), respectively. The independent variable is GV calculated as the total number of grammatical errors in MD&A (Management Discussion & Analysis) scaled by the number of words in the MD&A. FOG is the number of words per sentence plus the percentage of words that are complex (i.e., having three or more syllables). This sum is scaled by a constant (0.4) such that the Fog value approximates the number of years of formal education required to understand the text (see Li [2008]). SIZE denotes firm size defined as the log of total assets. MTB is the market-to-book ratio defined as the ratio of market value of equity plus the book value of debt scaled by the book value of assets. OPERATINGCYLE is a measure of the operating cycle of the firm. SGR is the change in sales from year t and t-1 scaled by sales in year t. LEV is the book debt scaled by total assets multiplied to 100. Loss is 1 if the firm's current earnings Income Before Extraordinary is negative, 0 otherwise. CFO/AT is the ratio of CFO to sales and ROA is return on assets, defined as income before extraordinary items scaled by total assets. Robust standard errors clustered by firm x year are presented in parentheses under the coefficients. We include year x industry fixed effects (defined as the 12 industry Fama & French code level). (***), (**), (*) denote statistical significance at the 1%, 5% and 10% levels, respectively.

Model: Discretionary Accruals$_t = \beta_0 + \boldsymbol{\beta_1 GV_t} + \beta_2$*Firm Controls* $+ \delta + \zeta$

	Absolute DCA	
	Column (1)	Column (2)
VARIABLES	ABS JONES$_t$	ABS MOD JONES$_t$
GV_t	0.036**	0.046**
	(0.017)	(0.019)
FOG_t	-0.002	-0.002
	(0.002)	(0.002)
$SIZE_t$	-0.016***	-0.018***
	(0.003)	(0.003)
MTB_t	0.011***	0.016***
	(0.002)	(0.003)
$OPERATINGCYCLE_t$	-0.026***	-0.026***
	(0.006)	(0.008)
SGR_t	-0.037**	-0.031*
	(0.014)	(0.017)
LEV_t	0.067***	0.075***
	(0.018)	(0.018)
$LOSS_t$	-0.031**	-0.028**
	(0.012)	(0.013)
CFO/AT_t	0.235***	0.292***
	(0.034)	(0.036)
ROA_t	-0.281***	-0.303***
	(0.034)	(0.036)
Constant	0.516***	0.547***
	(0.047)	(0.054)
Observations	42,044	42,015
Ind x Year FE	Yes	Yes
Adj. R-squared	0.364	0.396

IA.3 CFO Changes and Grammatical Violations

We provide an alternative measure of firms' investment in accounting resources using executive compensation. One important investment in accounting resources is the quality of the CFO. Presumably, more qualified and experienced CFOs will earn larger salaries. Using Execucomp, we measure the percentage increase in each CFO's salary per year. We adjust this percentage change by industry and size, using the Fama and French 12 industry classifications and market capitalization deciles. Observations without a CFO name in Execucomp are excluded from the sample. We examine whether firms that switch to a more qualified CFO subsequently report fewer grammatical violations (GVs) in their financial statements.

We present the results in Table IA.3.[2] The interaction term, SWITCH_CFO x △COMP_CFO_ADJ is negative and significant. The interpretation of our results is that grammatical violations decline when a firm changes CFOs and the increase in the new CFO's salary is greater than similarly sized firms within the same industry. In other words, when firms decide to improve the quality of their CFO, generating a large positive change in the salary of the new CFO relative to the old CFO, there is a marked decline in grammatical violations subsequent to the change. Thus, our results using changes in CFO salaries at the time of a CFO switch support our conclusion that GVs capture firms' investments in accounting resources.

[2] We note that the main effect, △COMP_CFO_ADJ, is not significant. This may be explained by a number of factors. For example, CFO salaries tend to increase over time. We isolate changes in pay at the time of CFO switches to provide better identification of greater investments in accounting resources.

Table IA.3: CFO Changes and Grammatical Violations

The dependent variable GV is calculated as the total number of grammatical errors in MD&A scaled by the number of words in the MD&A. SWITCH_CFO is an indicator variable equal to 1 if there is a CFO switch and 0 otherwise. △COMP_CFO_ADJ is calculated as the yearly change of the sum of CFO salary and bonus adjusted by 12 industry Fama & French code level and size. AFTER_HTML is an indicator variable equal to one when the issuance of SEC report is disclosed using HTML technology or zero if ASCII. MARKETCAP denotes firm market value. σ(CFO) is the standard deviation of CFO. Dividend is an indicator variable that takes the value of one if the firm paid a dividend. σ(EARN) is the standard deviation of earnings. OPERATINGCYLE is a measure of the operating cycle of the firm. FIRM_AGE is the number of years the firm has Compustat data. NBSEG is the logarithm of 1 plus the number of business segments. LEV is the book debt scaled by total assets multiplied to 100. Loss is an indicator variable that takes the value of one if net income before extraordinary items is negative, zero otherwise. Robust standard errors clustered by firm x year are presented in parentheses under the coefficients. We include year x industry fixed effects (defined as the 12 industry Fama & French code level) in column (I) and year and firm fixed effect in column (II). (***), (**), (*) denote statistical significance at the 1%, 5% and 10% levels, respectively.

Table IA.3: CFO Changes and Grammatical Violations

Model: $GV_t = \beta_0 + \beta_1 SWITCH\_CFO_t + \beta_2 \triangle COMP\_CFO\_ADJ_t + \beta_3$ SWITCH_CFO x $\triangle$COMP_CFO_ADJ + $\beta_4 Firm\ Controls + \delta + \zeta$

VARIABLES	(1) GV_t
$SWITCH\_CFO_t$	-0.011
	(0.010)
$\triangle COMP\_CFO\_ADJ_t$	0.003
	(0.002)
$SWITCH\_CFO_t$ x $\triangle COMP\_CFO\_ADJ_t$	-0.008**
	(0.003)
$AFTER\_HTML_t$	-0.011
	(0.034)
$MARKETCAP_t$	-0.006
	(0.010)
$STDDEVCFO_t$	-0.005
	(0.060)
$DIVIDEND_t$	-0.000
	(0.013)
$EARN\_VAR_t$	0.013
	(0.015)
$OPERATINGCYCLE_t$	-0.001
	(0.013)
$FIRM\_AGE_t$	-0.001
	(0.048)
$NBSEG_t$	-0.040**
	(0.019)
LEV_t	-0.001
	(0.002)
$LOSS_t$	-0.007
	(0.010)
Constant	0.300*
	(0.153)
Observations	2,402
Firm FE	Yes
Year FE	Yes
Adj. R-squared	0.481

IA.4 Additional Descriptive Statistics: HTML Adoption

We use the adoption of HTML as evidence of investments in accounting resources. We provide an example of a balance sheet disclosed by Apple in the year of 2000 using ASCII technology in Figure IA.1 and another example of a balance sheet disclosed by Apple in HTML after the adoption in the year of 2001 in Figure IA.2. To provide additional information on adoptions over time, we present the results in Table IA.4. In Panel A, the highest percentage of firms adopt HTML in 2002, 2001, and 2003, respectively. By the end of 2012, almost all firms had adopted HTML. In Panel B, the majority of firms adopt HTML in the fourth quarter.

With respect to the mandated adoption of XBRL in June 2009, firms can comply with XBRL using HTML or ASCII formats. Thus, the adoption of XBRL is not a substitute to HTML, although arguably a complement. We note that XBRL adoption has a similar effect on grammatical violations. Using the ratio of XBRL tags by the number of Compustat tags as a measure for the level of XBRL adoption, we find that grammatical violations decrease with the level of XBRL adoption (*untabulated*).

Figure IA.1: Balance sheet disclosed using ASCII technology

```
<PAGE>
                        CONSOLIDATED BALANCE SHEETS

                    (IN MILLIONS, EXCEPT SHARE AMOUNTS)

<TABLE>
<CAPTION>
                                                    SEPTEMBER 30, 2000   SEPTEMBER 25, 1999
                                                    ------------------   ------------------
<S>                                                 <C>                  <C>
                                    ASSETS:

Current assets:
  Cash and cash equivalents.....................          $1,191               $1,326
  Short-term investments........................           2,836                1,900
  Accounts receivable, less allowances of $64 and $68,
    respectively................................             953                  681
  Inventories...................................              33                   20
  Deferred tax assets...........................             162                  143
  Other current assets..........................             252                  215
                                                          ------               ------
    Total current assets........................           5,427                4,285
  Property, plant, and equipment, net...........             313                  318
  Non-current debt and equity investments.......             786                  339
  Other assets..................................             277                  219
                                                          ------               ------
    Total assets................................          $6,803               $5,161
                                                          ======               ======
```

Figure IA.2: Balance sheet disclosed using HTML technology

CONSOLIDATED BALANCE SHEETS

(In millions, except share amounts)

	September 29, 2001	September 30, 2000
ASSETS:		
Current assets:		
Cash and cash equivalents	$ 2,310	$ 1,191
Short-term investments	2,026	2,836
Accounts receivable, less allowances of $51 and $64, respectively	466	953
Inventories	11	33
Deferred tax assets	169	162
Other current assets	161	252
Total current assets	5,143	5,427
Property, plant, and equipment, net	564	419
Non-current debt and equity investments	128	786
Other assets	186	171
Total assets	$ 6,021	$ 6,803

Table IA.4: Descriptive Statistics Regarding HTML Adoption

Panel A: Year of HTML adoption

Adoption Year	Freq.	Percent of Adoption	Cum.
1999	1,723	3.04%	3.04
2000	3,579	6.32%	9.37
2001	10,568	18.68%	28.04
2002	12,349	21.82%	49.87
2003	7,962	14.07%	63.94
2004	5,538	9.79%	73.72
2005	3,992	7.05%	80.78
2006	3,618	6.39%	87.17
2007	2,676	4.73%	91.9
2008	2,158	3.81%	95.71
2009	1,028	1.82%	97.53
2010	680	1.2%	98.73
2011	584	1.03%	99.76
2012	88	0.16%	99.92
2013	7	0.01%	99.93
2014	6	0.01%	99.94
2015	1	0%	99.95
2016	8	0.01%	99.96
2017	22	0.04%	100
2018	1	0%	100
Total	56,588	100	

Panel B: Quarter regarding HTML adoption

Quarter adoption	Freq.	Percent
1	1,707	3.02%
2	2,410	4.26%
3	23,797	42.05%
4	28,674	50.68%
Total	56,588	100

IA.5 Robustness Check for XBRL and Accounting Disaggregation

Our study is similar to recent work that examines financial reporting complexity using XBRL tags (Hoitash and Hoitash [2018]) and disclosure quality using the level of disaggregation in accounting disclosures (Chen et al. [2015]). Our measure differs conceptually by capturing the innate *ability* of the accounting system to capture operating performance, which differs from manager's disclosure incentives related to reporting complexity or disaggregation. Empirically, we demonstrate that our main results remain robust to including both of these measures in the regression. Thus, we demonstrate that GVs capture a conceptually and empirically unique measure.

We begin by examining the relation between GVs and accounting reporting complexity (ARC) based on the count of accounting items (XBRL tags) disclosed in 10-K filings as proposed in Hoitash and Hoitash [2018]. Given that more complex financial reporting is likely to be associated with more grammatical violations, we expect a positive correlation between GV and ARC. Contrary to our expectation, ARC is negatively correlated with GV (*untabulated*). In a robustness test, the ratio of XBRL tags divided by the number of non-missing Compustat items is also negatively correlated with grammatical violations (*untabulated*).

In Table IA.5 column (1), we include ARC in the regression model that predicts the likelihood of a future restatement due to unintentional errors. As expected, accounting reporting complexity, the natural log of ARC, is positively correlated with the likelihood of a future restatement. However, grammatical violations (GVs) remain positive and significant. Combined, these findings suggest that GV and ARC depict different constructs.

We also examine the relation between GVs and disaggregation quality (DQ) following Chen et al. [2015]. Unlike ARC, we do not find any association between GVs and DQ in univariate tests (*untabulated*). We control for DQ in our main analysis predicting future restatements due to errors in Table IA.5 column (2). As expected, DQ is negative and significant, suggesting that greater disclosure quality is negatively correlated with future restatements. In addition, GVs remain positive and significant after including DQ in the regression. Collectively, the results support our argument that GVs capture a different underlying construct: accounting resources.

Table IA.5: Restatement Prediction Model and the Role of Grammatical Violations with XBRL and Accounting Disaggregation Controls

The dependent variable is RESTATE_ERRORS that denotes a restatement regarding accounting or clerical errors anytime in two years after the grammatical error. The independent variable is GV, calculated as the total number of grammatical violations in the MD&A scaled by the number of words in the MD&A. LNARC is the log of is the count of unique XBRL tags per financial statement disclosure as defined in Hoitash and Hoitash [2018] plus one. *DQ* is the count of non-missing accounting data items to measure accounting disaggregation as defined by (Chen, Miao, and Shevlin [2015]). FOG is the measure for readability as in Li [2008]. ROA is income before extraordinary items. *RET* denotes annual (excess) return. NCAP is an indicator equal to one when the firm issued debt or equity, and 0 otherwise. LEV denotes book leverage. MTB is the market-to-book ratio. *SGR* is the growth in sales over the previous year. NBSEG is the logarithm of 1 plus the number of business segments. RETVOL denotes stock return volatility. MARKETCAP is the log of market value. Robust standard errors clustered by firm x year are presented in parentheses under the coefficients. We include year and industry fixed effects (defined as the 2-digit SIC industry classification code). (***), (**), (*) denote statistical significance at the 1%, 5% and 10% levels, respectively.

Model: $Restatement_{t+i} = \beta_0 + \boldsymbol{\beta_1 GV_t} + \boldsymbol{\beta_2 LNARC_t} + \boldsymbol{\beta_3 DQ_t} + \beta_4 Firm\ Controls + \delta + \zeta$

VARIABLES	Column (1) $RESTATE\_ERRORS_{t+i}$	Column (2) $RESTATE\_ERRORS_{t+i}$
GV_t	0.024**	0.033**
	(0.012)	(0.015)
$LNARC_t$	0.057***	
	(0.008)	
DQ_t		-0.101*
		(0.061)
FOG_t	0.000	-0.001
	(0.001)	(0.002)
ROA_t	0.006	-0.006
	(0.006)	(0.013)
RET_t	0.005	-0.002
	(0.005)	(0.006)
$NCAP_t$	0.003	0.021***
	(0.006)	(0.007)
LEV_t	0.001	0.003**
	(0.001)	(0.002)
MTB_t	-0.269***	-0.414**
	(0.102)	(0.191)
SGR_t	0.019***	0.012*
	(0.004)	(0.007)
$NBSEG_t$	0.005	0.007
	(0.009)	(0.012)
$RETVOL_t$	0.014	0.163***
	(0.036)	(0.045)
$MARKETCAP_t$	-0.005***	-0.001
	(0.002)	(0.002)
Constant	-0.200***	0.204***
	(0.044)	(0.056)
Observations	14,172	13,075
Ind and Year FE	Yes	Yes
Adj. R-squared	0.0369	0.0214

IA.6 Rules Retained and Removed in Calculating Grammatical Violations

We calculate grammatical violations in financial statements using the *LanguageTool* project. We run the program on the initial (broader) sample of 188,323 MD&As and document a total of 370 rules as identified by *LanguageTool*.[3] We modify the list of grammatical rules in the program to adapt these generic grammar rules to the language used in financial reports. To accomplish this, we reviewed the grammatical rules by hand and eliminated 273 rules that generated a large number of Type II errors and kept 99 rules that reliably capture grammatical errors. The list of rules eliminated and retained, including an example and the rationale behind each rule, is provided in the Excel spreadsheet attached to this appendix.

[3] See http://wiki.languagetool.org/.

Variable Definitions Used in the Internet Appendix

ABS JONES	The absolute discretionary accruals measured using the Jones models.
ABS MOD JONES	The absolute discretionary accruals measured using the modified Jones models.
DQ	The count of non-missing accounting data items to measure accounting disaggregation as defined by (Chen, Miao, and Shevlin 2015).
LNARC	Log of is the count of unique XBRL tags per financial statement disclosure as defined in Hoitash and Hoitash (2018) plus one
△COMP_CFO_ADJ	CFO compensation yearly change calculated as the yearly change of the sum of CFO salary and bonus adjusted by 12 industry Fama & French code level and size.
SWITCH_CFO	Indicator variable equal to 1 if there is a CFO switch and 0 otherwise