Appendix A

Appendix A1

Association between *IRQ_TA* and EY ratings.

This appendix considers the association between *IRQ_TA* and the EY rating of integrated reporting quality used by Barth et al. (2017). In brief, each year EY South Africa publicly announces rankings of the quality of the integrated reports issued by the top 100 JSE listed firms based on market capitalization. The quality rankings are based on the Integrated Reporting Framework's (Framework) fundamental concepts, guiding principles, and content elements, which means a firm not delivering on those characteristics receives a low rank. Integrated report quality is assessed by three independent, experienced professors at the University of Cape Town who have been on the adjudication panel since the inception of the EY awards. Unlike *IRQ_TA* which uses objective measures, the EY ratings incorporate human judgement.

We convert the EY ratings into percentile ranks, where *IRQ_EY* is (firm rank – 1) / (number of firms – 1) as in Barth et al. (2017). We first examine the relation between *IRQ_EY* and the seven individual textual attributes one at a time because the textual attributes are highly correlated. We find that *IRQ_EY* is significantly associated with all attributes. In particular, the results suggest that higher *IRQ_EY* is associated with integrated reports with less boilerplate language, more readability, more numeric disclosure, more words, and fewer redundant and sticky words. The only exception is the coefficient on specificity, which is insignificant.

Next, we regress *IRQ_EY* on *IRQ_TA* and the control variables and find that the EY ranking also is significantly and positively associated with *IRQ_TA.* These analyses help validate our text-based IRQ measure because *IRQ_TA* is correlated with the EY ratings based on expert opinions. This provides support for the use of *IRQ_TA* as a measure of IRQ, and more generally, sustainability information, particularly in large samples where expert ratings are not available.

We further test how our text-based IRQ measures link to the Framework's guiding principles and content elements. Untabulated statistics reveal that *IRQ_TA* is significantly correlated with all components of the EY rankings and that the correlation coefficients for three—strategic focus and future orientation, connectivity of information, and organizational overview and external environment—are greater than 0.5. In terms of individual textual attributes, reports with less boilerplate language, more readability, more numeric disclosure, more words, and fewer redundant and sticky words are correlated with higher quality in EY-based IRQ components. In addition, our three IRQ factors are significantly correlated with the overall EY ranking and the components. In general, the analysis in this Appendix reveals a high correlation between the EY scores and our text-based measures, which provides external validity to our measures.

Appendix A2
Identification of integrated reporting topics using PhraseLDA.

Several prior studies in accounting and finance use Latent Dirichlet Allocation (LDA), e.g., Ball, Hoberg, and Maksimovic (2015), Dyer et al. (2017), Hoberg and Lewis (2017), and Huang, Lehavy, Zang, and Zheng (2017). LDA is an unsupervised topic modelling technique based on a Bayesian probabilistic model that was developed by Blei, Ng, and Jordan (2003). It assumes that each document is a mixture of various latent topics, and each latent topic is a distribution of words. By examining the co-occurrence of words in documents, LDA can uncover the underlying latent topics and estimate the probability that a particular sentence belongs to each of these latent topics.

However, the traditional LDA model has many drawbacks, which motivates us to use a more advanced topic modeling method. In particular, we use PhraseLDA (El-Kishky et al., 2014) to discover the latent topics of a collection of integrated reports. PhraseLDA uses an efficient phrase mining technique and a statistical significance measure to segment the text while simultaneously removing false phrases. In the PhraseLDA graph model, a phrase is represented by a clique, and a potential function is used to coerce all variables in the clique, i.e., words in the phrase, to take on the same latent topic.

The advantages of PhraseLDA over the traditional LDA can be summarized as follows. (1) Traditional LDA uses a list of unigram words to describe individual topics, which makes it difficult to interpret the themes of individual topics. In contrast, a list of phrases identified by PhraseLDA provides a more interpretable and accurate description of the theme of a topic. (2) The unigram words used by traditional LDA may be ambiguous. For example, the word "model" can mean a fashion model or a scientific or mathematical model such as a "topic model." The use of phrases by PhraseLDA helps avoid this ambiguity. (3) Traditional LDA is built upon a 'bag-of-words' and the order of words is ignored. As a result, when inferring the topic assignment for a word, the topic of a far-away neighbor word in the same document has the same impact as a near-by word. PhraseLDA captures a stronger correlation between the words in a phrase than that expressed by LDA.

To implement PhraseLDA topic modeling, we need to specify how many topics PhraseLDA should identify. To gain insight into the extent to which identifying more topics results in more—or less—interpretable topics, we implement the topic modeling with 100, 125, and 150 topics. We use the coherence and perplexity metrics to measure the quality of topics (Newman, Grieser, & Baldwin, 2010; Wallach, Murray, Salakhutdinov, & Mimno, 2009). For both metrics, we find that increasing the number of topics beyond 100 does not improve the coherence or perplexity metrics. Based on the lists of most probable phrases of individual topics, we manually group the 100 PhraseLDA topics into integrated report categories. This enables us to present details more concisely and investigate whether the association between IRQ and synchronicity varies by the category of the disclosure.

We identify integrated report categories based on the eight content elements outlined in the Integrated Reporting Framework (Framework). We manually categorize the 100 topics into a unique integrated report category based on the majority of the phrases in each topic.[1] Also, because a substantial portion of sentences in integrated reports relate to disclosures around financial statements and International Financial Reporting Standards (IFRS) compliance, we include IFRS as an additional category.

[1] For example, topics on board, committee, meetings, responsibilities are categorised as governance, and topics on revenue growth, earnings, dividends are categorised as performance.

We calculate the textual attributes of each category based on the topics included in that category and the sentences allocated to each topic by the PhraseLDA model. For some categories, although the PhraseLDA model discovers related topics, there are only a few sentences of the related topics in most integrated reports. Because of the limited coverage of these topics—and consequently their assigned categories—in the sentences, we are able to generate outcomes for five categories: governance, performance, risks and opportunities, IFRS, and other. The other category includes low-covered sentences related to the following Framework elements and guiding principles: organizational overview and external environment, business model, strategy and resource allocation, outlook, basis of preparation and presentation, and general reporting guidance.

Appendix A3
Sensitivity analyses.

1. Changes in IRQ and synchronicity

Equation (1) specifies the relation between the level of synchronicity and integrated report quality, which could be affected by time-invariant correlated omitted variables. To address this type of endogeneity, we follow and Grewal et al. (2021) Gul et al. (2011) and estimate equation (1) replacing *IRQ_TA* with an indicator variable, *Post*, which is constructed to identify changes in IRQ. To construct *Post*, we identify each firm's first change in *IRQ_TA* rank as well as any subsequent changes in rank that are at least three years after the previous change.[2] We identify the year of each of these rank changes as year t and retain observations for years $t-1$, t, and $t+1$. We set *Post* equal to one for firm years t and $t+1$ and to zero for year $t-1$. We then create subsamples for positive and negative changes in rank in year t and estimate the equation separately for each subsample. If changes in *IRQ_TA* ranks reflect decreases (increases) in firm-specific information in integrated reports, we predict the coefficient on *Post* is negative (positive) in the positive (negative) *IRQ_TA* rank changes subsample.

To ensure that changes in *IRQ_TA* ranks are not affected by changes in sample composition, we estimate the equation based on the 47 firms with *IRQ_TA* in all seven years. In addition, we re-measure IRQ by re-ranking *IRQ_TA* for these firms across sample years. The resulting sample is 252 observations for 47 firms. We also conduct this analysis including only large changes in IRQ ranks, where large changes are those in the upper and lower quartiles of the IRQ rank change distribution pooled

[2] Although Grewal et al. (2021) include only the first change, the study notes that its findings are insensitive to including multiple changes. We include multiple changes to increase sample size. However, even with multiple changes our sample size is small and, thus, the results should be interpreted with caution.

across years. The resulting sample is 165 observations for 35 firms.

Appendix A Table 1 presents the findings. The first (second) set of columns presents findings for firms with one or more (large) *IRQ_TA* changes, separately for positive and negative *IRQ_TA* changes. In column (1) for positive *IRQ_TA* changes, the *Post* coefficient is significantly negative, which reveals, consistent with H1, that increases in integrated report quality are associated with decreases in synchronicity. For negative *IRQ_TA* changes in column (2), the *Post* coefficient is insignificant. Consistently, for large positive (negative) *IRQ_TA* changes, the *Post* coefficient is significantly (insignificantly) negative as shown in column (3) ((4)). Taken together, the findings in Online Appendix Table 1 reveal that positive changes in integrated report quality are associated with decreases in synchronicity.

2. Alternative IRQ proxy

We use the quality of the external assurance of integrated reports as an alternative proxy for IRQ. The quality of the external assurance (*EAQ*) is measured as the average score based on four binary variables: (1) external assurance versus no external assurance, (2) accounting versus non-accounting assurance provider, (3) Big 4 versus non-Big 4 accounting assurance provider, (4) reasonable/high versus limited/moderate assurance level, where external assurance provided by a Big4 accounting firm at a reasonable/high level is considered the highest quality.[3] We reestimate our analyses presented in Tables 4 and 5 by replacing *IRQ_TA* with *EAQ*.

Untabulated findings reveal that, in the revised equation (1), the coefficient on *EAQ* is negative and marginally significant (coef. = –0.197; *t*-stat. = –1.52). When we reestimate equation (2) with *EAQ*, the main effect of *EAQ* is negative and significant

[3] We obtain data for this test from Thomson Reuters ASSET4, the GRI Report List, hand collection from integrated reports and/or other extended external reports, and the independent external assurance statement on the firm’s website.

regardless of whether *HHI* or *4Firm* is the proxy for proprietary costs (coefs. = –0.369 and –0.330; *t*-stat. = –2.40 and –2.24). The coefficient on the interaction between *EAQ* and *Prop* is positive and marginally significant when *HHI* is the proprietary costs proxy (coef. = 0.253; *t*-stat. = 1.29) and insignificant when *4Firm* is the proxy (coef. = 0.189; *t*-stat. = 0.86). Although weaker, these results are consistent with our main analysis for H1 and H2.

3. Alternative proprietary cost proxies

We compute three alternative measures of proprietary costs related to growth opportunities and intangible assets: the market-to-book ratio (the number of common shares outstanding multiplied by the end-of-year stock price divided by the book value of common equity), which relates to firms' growth opportunities (Ajinkya, Bhojraj, & Sengupta, 2005; Bamber, Jiang, & Wang, 2010); recognized intangible assets net of goodwill, scaled by total assets; and whether the firm is a member of one of the seven Fama and French 48 industries in which the redaction of contract information in mandatory disclosures in the U.S. is most prevalent (Heinle, Samuels, & Taylor, 2023). Redaction of mandatory disclosures reflects concern over revealing proprietary information, which is consistent with these industries being characterized by high intangible assets. We convert the market-to-book ratio and intangible assets to indicator variables that equal one for values above the median of the full sample distribution and zero otherwise. The prevalent redaction variable is an indicator variable that equals one if the if the Fama and French 48 industry classification is pharmaceutical products, business services, electronic equipment, transportation, medical equipment, communication, or retail, and zero otherwise.

Untabulated statistics from reestimating equation (2) with interactions between these proprietary cost measures and IRQ, reveal that none of the interaction coefficients is significant. These findings indicate that, in our context, proprietary costs related to product market competition are more relevant than proprietary costs related to growth opportunities and intangible assets.

Additional references

Ajinkya, B., Bhojraj, S., & Sengupta, P. (2005). The association between outside directors, institutional investors and the properties of management earnings forecasts. *Journal of Accounting Research*, *43*(3), 343–376.

Ball, C., Hoberg, G., & Maksimovic, V. (2015). Disclosure, business change and earnings quality. *Available at SSRN 2260371*

Bamber, L.S., Jiang, J., & Wang, I.Y. (2010). What's my style? The influence of top managers on voluntary corporate financial disclosure. *The Accounting Review*, *85*(4), 1131–1162.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Heinle, M., Samuels, D., & Taylor, D. (2023). Disclosure substitution. *Management Science, 69*(8), 4774–4789.

Hoberg, G., & Lewis, C. (2017). Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance, 43*, 58–85.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2017). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science, 64*(6), 2833–2855.

Newman, D., La, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 100–108.

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, 1105–1112.

Appendix A Table 1
Changes in integrated report quality and synchronicity.

	Changes		Large Changes	
	Positive	Negative	Positive	Negative
Variable	(1)	(2)	(3)	(4)
Post	−0.3836**	−0.0701	−0.6150***	−0.0374
	(−1.75)	(−0.75)	(−2.47)	(−0.23)
Size	0.1599	0.1985**	0.1948	0.2627**
	(0.96)	(2.19)	(1.16)	(2.57)
II_Net	−0.0063	−0.0019	−0.0083	−0.0061
	(−0.98)	(−0.35)	(−0.73)	(−1.49)
MTB	0.0333	0.0675*	0.0831	0.0318
	(0.80)	(1.98)	(1.19)	(0.61)
Earn_Vol	0.0779	−10.6349***	1.3457	−7.8088*
	(0.02)	(−3.49)	(0.23)	(−1.68)
Trade_Vol	−0.0354	−0.1368	−0.0375	−0.0810
	(−0.40)	(−1.35)	(−0.37)	(−0.63)
LowAQ	0.6332	−0.2224	0.3671	−0.1509
	(1.04)	(−1.04)	(0.64)	(−0.70)
FogPR	−0.1022*	−0.1013**	−0.0654	−0.0281
	(−1.96)	(−2.24)	(−1.11)	(−0.55)
CSR_SA	−0.0128	−0.1495	−0.0603	−0.0861
	(−0.08)	(−0.93)	(−0.31)	(−0.41)
Prime	-	−0.4687**	-	−0.0373
		(−2.10)		(−0.11)
$MBeta1_{t+1}$	0.7371***	0.8004***	0.8295***	0.6823***
	(11.97)	(20.82)	(6.56)	(8.11)
$MBeta2_{t+1}$	0.0373	0.0048	0.0089	0.0131
	(1.10)	(0.29)	(0.19)	(0.75)

Variable	Changes Positive (1)	 Negative (2)	Large Changes Positive (3)	 Negative (4)
Adjusted R^2	0.858	0.862	0.888	0.927
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
N	123	129	72	93

Appendix A Table 1 presents summary statistics from estimating an equation of synchronicity on *Post* and control variables with *t*-statistics in parentheses. Changes (Large changes) are the firm's first change in *IRQ_TA* (in the upper and lower quartiles of the distribution) and any subsequent changes at least two years after the previous change. We retain firm years, i.e., $t-1$, t, and $t+1$, where t is the change year. Positive (negative) refers to positive (negative) changes in year t. See Appendix 1 for variable definitions. Standard errors are clustered by firm and by year based on bootstrapping methods using 10,000 iterations. *, **, and *** denote significance at a 10%, 5%, and 1% level based on a two-tailed test, except for *Post*, which is based on a one-tailed test. Sample of firm-year observations from 2012 to 2017.