ÁREA TEMÁTICA: GESTÃO SOCIOAMBIENTAL
CATEGORIA: ENANGRAD PLENO

CHECKLIST ESG PARA EVIDENCIAÇÃO DOCUMENTAL: DESENVOLVIMENTO E APLICAÇÃO-PILOTO EM COMPANHIAS LISTADAS NA B3

Resumo

Este estudo desenvolve e examina, em aplicação-piloto, um checklist para diagnóstico inicial da evidenciação ambiental, social e de governança (ESG) em documentos públicos corporativos. O instrumento foi construído a partir de referenciais como GRI Standards, ISO 26000, ABNT PR 2030-1:2024 e metodologia de scores da LSEG. A versão aplicada contém 15 itens, cinco por dimensão, codificados como Sim = 1, Parcial = 0,5 e Não ou não disponível = 0. A amostra intencional reúne 41 companhias listadas na B3, distribuídas em cinco setores, totalizando 615 classificações. A média global foi 77,6%, a mediana 86,7% e o desvio-padrão 25,8 pontos percentuais. A dimensão Social apresentou a maior média, 87,3%, seguida por Governança, 73,7%, e Ambiental, 71,7%. Os menores resultados ocorreram em meta climática quantitativa, asseguração externa e remuneração vinculada a metas ESG. O alfa de Cronbach foi 0,934 no conjunto, mas a forte associação entre pontuação e disponibilidade documental exige interpretação cautelosa. Conclui-se que o checklist é operacional para triagem e organização de evidências, porém ainda não constitui instrumento validado nem medida de desempenho ESG. Sua consolidação requer protocolo documental uniforme, critérios mais específicos, avaliação por especialistas e teste de concordância entre avaliadores.

Palavras-chave: ESG; evidenciação; checklist; sustentabilidade corporativa; B3.

Abstract

This study develops and examines, through a pilot application, a checklist for the initial diagnosis of environmental, social, and governance (ESG) disclosure in public corporate documents. The instrument was based on references such as the GRI Standards, ISO 26000, ABNT PR 2030-1:2024, and the LSEG score methodology. The applied version contains 15 items, five per dimension, coded as Yes = 1, Partial = 0.5, and No or not available = 0. The purposive sample comprises 41 companies listed on B3 across five sectors, totaling 615 classifications. The global mean was 77.6%, the median was 86.7%, and the standard deviation was 25.8 percentage points. The Social dimension had the highest mean, 87.3%, followed by Governance, 73.7%, and Environmental, 71.7%. The lowest results concerned quantitative climate targets, external assurance, and compensation linked to ESG targets. Cronbach's alpha was 0.934 for the full set, but the strong association between scores and document availability requires caution. The checklist is operational for screening and organizing evidence, but it is not yet a validated instrument or a measure of ESG performance. Consolidation requires a uniform document protocol, more specific criteria, expert assessment, and inter-rater agreement testing.

Keywords: ESG; disclosure; checklist; corporate sustainability; B3.

1. Introdução

A agenda ambiental, social e de governança ampliou o conjunto de informações utilizado para avaliar organizações. Emissões, uso de recursos, condições de trabalho, diversidade, integridade e estrutura decisória passaram a integrar relatórios corporativos e decisões de investidores, reguladores e demais

partes interessadas. Esse movimento aumentou a demanda por informação não financeira comparável, material e verificável, mas também evidenciou diferenças de escopo, profundidade e linguagem entre os documentos divulgados pelas empresas (Global Reporting Initiative, 2021; OECD, 2020).

A multiplicidade de padrões não produz, por si só, avaliações equivalentes. Os resultados de provedores de ratings podem divergir porque cada método define de forma própria os temas avaliados, os indicadores, os pesos e o tratamento de informações ausentes. Berg, Kölbel e Rigobon (2022) mostram que diferenças de escopo, mensuração e ponderação explicam parcela relevante da divergência entre ratings ESG. Por isso, qualquer instrumento sintético deve explicitar o que mede, de onde vêm seus critérios e quais limites impedem que a pontuação seja interpretada como verdade abrangente sobre a sustentabilidade de uma organização.

No cotidiano empresarial, instrumentos extensos podem ser pouco acessíveis para um diagnóstico inicial. Referenciais como os GRI Standards, a ISO 26000, a ABNT PR 2030-1:2024 e a metodologia da LSEG oferecem bases conceituais e operacionais robustas, porém demandam seleção, contextualização e capacidade documental. Um checklist pode funcionar como porta de entrada ao organizar perguntas básicas, orientar a busca de evidências e tornar visíveis lacunas de divulgação. Seu valor está menos em produzir um ranking definitivo e mais em indicar o que foi encontrado, o que precisa ser aprofundado e quais informações ainda não estão disponíveis.

Essa utilidade depende de uma delimitação essencial: presença documental não equivale à efetividade da prática. Uma política pública no portal corporativo comprova a divulgação daquele documento, mas não demonstra, sozinha, sua execução ou seus resultados. Da mesma forma, a ausência de uma evidência nas fontes consultadas não prova que a prática inexista. O instrumento examinado neste artigo deve, portanto, ser entendido como medida de evidenciação e organização documental, não como certificação ou rating de desempenho ESG.

Diante desse contexto, o problema de pesquisa é: como estruturar um checklist sintético, ancorado em referenciais reconhecidos, capaz de apoiar o diagnóstico inicial da evidenciação ESG em companhias brasileiras? O objetivo geral é desenvolver e examinar a operacionalidade de um checklist documental nas dimensões Ambiental, Social e de Governança. Especificamente, busca-se: sistematizar os critérios e sua rastreabilidade; aplicar o instrumento em uma amostra-piloto; descrever os padrões agregados de resposta; analisar sua consistência interna de forma exploratória; e identificar os ajustes necessários antes de uma validação formal.

A principal contribuição é metodológica. O estudo não propõe um novo rating de mercado, mas uma ferramenta preliminar, transparente e de baixo custo para organizar evidências públicas. A análise também contribui ao separar três conceitos frequentemente confundidos: preenchimento da planilha, disponibilidade documental e aderência ao critério. Essa distinção permite interpretar resultados sem transformar lacunas de informação em afirmações indevidas sobre desempenho corporativo.

O artigo está organizado em cinco seções. Após esta introdução, a fundamentação teórica discute materialidade, evidenciação e limites dos escores. A metodologia apresenta a construção do checklist, a amostra, a coleta e a codificação. A quarta seção analisa os resultados agregados e a consistência interna. Por fim, a conclusão reúne contribuições, limitações e etapas recomendadas para validação do instrumento.

2. Fundamentação teórica

2.1 ESG, materialidade e evidenciação

A evolução da sustentabilidade corporativa deslocou a avaliação empresarial para além do resultado financeiro de curto prazo. A perspectiva do Triple Bottom Line ajudou a consolidar a ideia de que resultados econômicos, ambientais e sociais precisam ser considerados conjuntamente (Elkington, 1997). Posteriormente, a agenda ESG aproximou esses temas das estruturas de risco, governança, alocação de capital e prestação de contas, com maior ênfase em métricas observáveis e informações úteis à decisão.

Materialidade é central nesse processo. Nos GRI Standards, a organização deve identificar e relatar seus impactos mais significativos sobre a economia, o meio ambiente e as pessoas (Global Reporting Initiative, 2021). Na lógica financeira, a materialidade também orienta a seleção de informações capazes de influenciar decisões de investidores. Embora as perspectivas não sejam idênticas, ambas rejeitam a ideia de que todos os temas possuam a mesma relevância em todos os setores e contextos.

A ISO 26000 organiza a responsabilidade social em temas como governança, direitos humanos, práticas de trabalho, meio ambiente, práticas leais de operação, consumidores e comunidade (ABNT, 2010). A ABNT PR 2030-1:2024, por sua vez, oferece conceitos, diretrizes e um modelo de avaliação e direcionamento ESG para organizações (ABNT, 2024). Esses referenciais contribuem para a construção de perguntas, mas não autorizam a redução automática de seus conteúdos a quinze indicadores equivalentes. A síntese exige reconhecer perdas de detalhe e finalidade diagnóstica limitada.

Metodologias de scores, como a da LSEG, ilustram o papel da divulgação corporativa e da materialidade setorial. A informação pública é transformada em métricas comparáveis por meio de regras de normalização e ponderação (LSEG, 2024). Esse processo é útil, mas depende da cobertura documental. Uma empresa com práticas pouco divulgadas pode receber avaliação inferior à de outra com maior capacidade de reporte, ainda que a efetividade operacional não siga a mesma ordem.

Por isso, a evidenciação deve ser tratada como objeto próprio. Ela reduz assimetrias informacionais, melhora a rastreabilidade e permite que interessados examinem políticas, metas e resultados. Contudo, qualidade de divulgação, desempenho substantivo e impacto não são sinônimos. Um checklist documental captura, em primeiro lugar, a presença e a clareza da evidência na fonte consultada. A interpretação correta preserva essa fronteira.

2.2 Checklists, comparabilidade e limites da síntese

Checklists são úteis quando transformam uma atividade complexa em sequência verificável de questões. Em diagnóstico ESG, podem apoiar triagem, padronização da leitura e registro de justificativas. Sua qualidade, entretanto, depende de definições operacionais claras. Itens muito amplos podem produzir respostas positivas por caminhos diferentes. Por exemplo, reunir comunidade, clientes ou impacto social em uma única pergunta amplia a chance de atendimento, mas dificulta saber qual aspecto foi efetivamente evidenciado.

A comparabilidade também exige regras uniformes de coleta. Empresas avaliadas com documentos de anos distintos, tipos documentais diferentes ou buscas de profundidade desigual não partem da mesma condição. O problema é particularmente sensível quando a ausência de um anexo é tratada como ausência

de divulgação pública. Em uma aplicação formal, a busca deve seguir o mesmo roteiro para todas as unidades: período de referência, fontes prioritárias, palavras de busca, registro do URL, data de acesso e procedimento para resolver documentos não encontrados.

Outro ponto é a codificação da ausência. A categoria 'Não' deveria ser reservada a situações em que a fonte permite concluir que o critério não é atendido. Quando a informação não é localizada ou não é suficiente para decisão, a classificação apropriada é 'N/D'. Misturar as duas situações não altera o escore quando ambas recebem zero, mas afeta a interpretação da cobertura documental e pode levar o leitor a confundir falta de evidência com evidência de inexistência.

A adoção de pontuação parcial reconhece que a evidenciação pode ser incompleta. Uma política sem indicador, uma meta sem prazo ou um programa sem monitoramento não equivalem ao atendimento integral, mas também não são idênticos à ausência. O valor intermediário facilita a síntese, desde que cada item possua critérios específicos para Sim, Parcial, Não e N/D. Sem essa rubrica, permanece espaço significativo para julgamento do avaliador.

Por fim, uma média não ponderada atribui o mesmo peso a temas distintos. Essa opção torna o instrumento simples, porém limita comparações entre setores. A literatura e os métodos de mercado indicam que a relevância dos temas varia por indústria (LSEG, 2024). Assim, um checklist genérico pode apoiar a organização inicial das evidências, mas não deve sustentar afirmações sobre qual setor ou empresa possui melhor desempenho ESG sem calibração de materialidade.

2.3 Consistência interna e validação

O alfa de Cronbach é uma medida de consistência interna baseada na relação entre os itens de uma escala (Cronbach, 1951). Valores elevados podem indicar que as respostas variam de maneira associada, mas não comprovam que os itens representem adequadamente o conteúdo pretendido. Também não demonstram estabilidade temporal, concordância entre avaliadores, validade externa ou capacidade preditiva (Nunnally; Bernstein, 1994; Tavakol; Dennick, 2011).

Essa cautela é ainda mais importante em instrumentos multidimensionais. Ambiental, Social e Governança não formam necessariamente um único construto refletivo, no qual todos os itens seriam manifestações intercambiáveis da mesma variável latente. O checklist pode funcionar como índice composto de temas diferentes. Nesse caso, o alfa é um diagnóstico exploratório da matriz de respostas, não um selo de validação.

A validação de conteúdo requer avaliação sistemática por especialistas, com análise de clareza, pertinência e representatividade dos itens. A confiabilidade da classificação documental deve ser examinada por dois ou mais avaliadores independentes, acompanhada por medida de concordância, como o kappa de Cohen. Aplicações em novas amostras e comparação com indicadores externos podem avaliar estabilidade e validade convergente. Essas etapas são posteriores ao teste de operacionalidade realizado neste estudo.

3. Metodologia

3.1 Delineamento e construção do instrumento

A pesquisa é aplicada, documental, descritiva e de abordagem quantitativa. Seu produto é um checklist para diagnóstico inicial de evidenciação ESG. O percurso compreendeu revisão dos referenciais, seleção de temas, formulação de itens, definição das categorias de resposta, aplicação-piloto e análise descritiva da

matriz resultante. O desenho não busca representar estatisticamente todas as companhias brasileiras, mas testar a operacionalidade do instrumento em ambiente com disponibilidade mínima de documentos públicos.

Os itens foram derivados de temas recorrentes nos GRI Standards, na ISO 26000, na ABNT PR 2030-1:2024, na metodologia da LSEG e em referenciais de governança. Para manter uma estrutura equilibrada, foram selecionados cinco critérios ambientais, cinco sociais e cinco de governança. O número reduzido favorece aplicação inicial, mas não reproduz integralmente nenhum dos referenciais de origem.

A Tabela 1 apresenta a versão aplicada. Os enunciados foram resumidos para caber no formato do artigo. Na etapa seguinte de desenvolvimento, itens que combinam alternativas amplas deverão ser divididos ou acompanhados de rubricas específicas. A rastreabilidade conceitual foi mantida por dimensão e fonte, enquanto a evidência empírica foi registrada em planilha com empresa, setor, item, status, justificativa, página e documento, quando disponíveis.

Tabela 1 - Estrutura do checklist aplicado

Item	Critério documental	Referenciais centrais
E1	Divulgação de emissões de gases de efeito estufa.	GRI; LSEG; CDP
E2	Meta climática quantitativa com prazo definido.	GRI; LSEG; CDP
E3	Consumo, captação ou reuso de água.	GRI; LSEG
E4	Resíduos, reciclagem ou economia circular.	GRI; LSEG; OECD
E5	Energia renovável ou eficiência energética.	GRI; LSEG
S1	Diversidade de gênero e/ou raça, especialmente em liderança.	ISO 26000; GRI; LSEG
S2	Indicadores de saúde e segurança.	ISO 26000; GRI; LSEG
S3	Treinamento, capacitação ou horas de treinamento.	ISO 26000; LSEG
S4	Direitos humanos e/ou gestão socioambiental de fornecedores.	ONU; GRI; LSEG
S5	Indicadores ou programas de comunidade, clientes ou impacto social.	ISO 26000; LSEG
G1	Composição do conselho e membros independentes.	IBGC; OECD; LSEG
G2	Comitê formal de auditoria, sustentabilidade ou ESG.	IBGC; OECD; LSEG
G3	Código de ética/anticorrupção e canal de denúncias.	ISO 26000; IBGC; LSEG
G4	Remuneração variável vinculada a metas ESG.	IBGC; LSEG

Item	Critério documental	Referenciais centrais
G5	Asseguração externa do relatório ESG/sustentabilidade.	GRI; OECD; LSEG

Fonte: elaboração própria a partir dos referenciais da pesquisa.

3.2 Amostra e coleta documental

A amostra intencional reuniu 41 companhias abertas listadas na B3: dez do setor de Construção, dez do setor Elétrico, oito do setor Industrial, sete de Petróleo/Energia e seis do setor Químico. A seleção foi definida antes da coleta e buscou heterogeneidade setorial e disponibilidade de informação. Como amostra não probabilística, ela permite examinar o instrumento, mas não autoriza generalização para o universo das empresas brasileiras.

A aplicação utilizou relatórios de sustentabilidade, relatos integrados, demonstrações e relatórios anuais, Formulários de Referência, estatutos, políticas, códigos corporativos, páginas de relações com investidores e bases públicas. Uma auditoria posterior da planilha mostrou que os documentos não pertenciam a um único ano-base: havia fontes entre 2020 e 2025, com profundidade variável entre empresas. Também foram encontradas referências a 'anexos enviados' ou 'anexos recebidos'. Por essa razão, a análise é apresentada como aplicação-piloto e os resultados por empresa e setor não são utilizados como ranking.

Ferramentas digitais e de inteligência artificial apoiaram a localização e a organização preliminar de trechos. Elas não foram consideradas fonte documental nem evidência autônoma. A classificação registrada na planilha permaneceu sob responsabilidade do pesquisador. Como não houve segundo avaliador independente, o estudo não estima concordância entre examinadores, o que constitui limitação relevante.

O protocolo ideal de replicação deve fixar um ano-base ou a regra 'documento mais recente disponível até uma data de corte', aplicar as mesmas fontes prioritárias, registrar título completo, URL e data de acesso e distinguir o limite organizacional da evidência. Esta última regra é importante para holdings e grupos empresariais, nos quais práticas de controladas ou investidas não devem ser atribuídas automaticamente à companhia analisada.

Tabela 2 - Composição da aplicação-piloto

Setor	Companhias	Classificações (15 por companhia)
Construção	10	150
Elétrico	10	150
Industrial	8	120
Petróleo/Energia	7	105
Químico	6	90
Total	41	615

Fonte: elaboração própria a partir da base da pesquisa.

3.3 Codificação, escores e análise

Cada item recebeu uma classificação categórica. 'Sim' foi codificado como 1 quando havia evidência clara de atendimento; 'Parcial', como 0,5 quando a evidência atendia apenas parte do critério; 'Não', como 0 quando a fonte permitia identificar não atendimento; e 'N/D', também como 0, quando a informação não foi localizada ou não era suficientemente clara. 'N/A' seria excluído do denominador, mas não ocorreu na aplicação.

Para a empresa i e o item j, a pontuação xij assumiu 1, 0,5 ou 0. O escore de cada dimensão foi calculado pela soma de seus cinco itens dividida por cinco e multiplicada por 100. O escore global correspondeu à soma dos 15 itens dividida por 15 e multiplicada por 100. Como todos os itens foram tratados como aplicáveis, os denominadores permaneceram fixos. Esse procedimento produz índice não ponderado de aderência documental e não incorpora materialidade setorial.

A análise utilizou frequências das categorias, média, mediana, desvio-padrão amostral, amplitude, médias por dimensão e por item. A completude formal foi definida como proporção de células preenchidas. A cobertura documental provisória correspondeu à proporção de classificações diferentes de N/D. A distinção entre 'Não' e 'N/D' não foi aplicada de maneira plenamente uniforme na base original; portanto, essa cobertura é informada apenas como diagnóstico da codificação inicial.

A consistência interna foi examinada com o alfa de Cronbach para os 15 itens e para cada bloco de cinco. Como verificação de sensibilidade, comparou-se a pontuação à disponibilidade documental e calculou-se o alfa de uma matriz binária de cobertura. Os resultados são exploratórios e não substituem avaliação de conteúdo, teste-reteste, análise fatorial ou concordância entre avaliadores. Adotou-se apresentação descritiva, sem testes de diferença entre setores, devido às limitações do protocolo documental.

4. Análise e discussão dos resultados

4.1 Distribuição das respostas e cobertura

A aplicação gerou 615 classificações. Foram registrados 445 'Sim' (72,4%), 64 'Parcial' (10,4%), 23 'Não' (3,7%) e 83 'N/D' (13,5%). Em conjunto, respostas positivas ou parciais representaram 82,8% da matriz. As categorias com pontuação zero somaram 17,2%. Esses percentuais descrevem a base codificada, mas a separação entre 'Não' e 'N/D' deve ser lida com cautela porque algumas justificativas de 'não localizado' foram registradas como 'Não'.

Todas as células da planilha foram preenchidas, resultando em 100% de completude formal. Isso não significa cobertura documental integral. Pela codificação original, 532 de 615 classificações foram diferentes de N/D, correspondendo a cobertura provisória de 86,5%. O valor deve ser revisado após padronização das categorias, pois uma célula preenchida com N/D registra uma decisão analítica, mas também indica que a evidência não estava disponível ou clara.

A distinção é substantiva. Um banco pode estar completo para análise e, ao mesmo tempo, conter lacunas importantes de divulgação. O resultado mais seguro desta aplicação não é afirmar que as empresas possuem 86,5% de cobertura ESG, mas que a busca original produziu evidência positiva ou parcial em 82,8% dos itens e manteve 17,2% com pontuação zero. O percentual restante depende da forma como se distingue ausência comprovada de ausência de informação.

Tabela 3 - Distribuição das respostas da aplicação-piloto

Categoria	Frequência	Participação	Pontuação
Sim	445	72,4%	1,0
Parcial	64	10,4%	0,5
Não	23	3,7%	0
N/D	83	13,5%	0
Total	615	100,0%	-

Nota: N/D = informação não disponível ou não suficientemente clara na fonte consultada.

Fonte: elaboração própria a partir da base da pesquisa.

4.2 Resultados globais e por dimensão

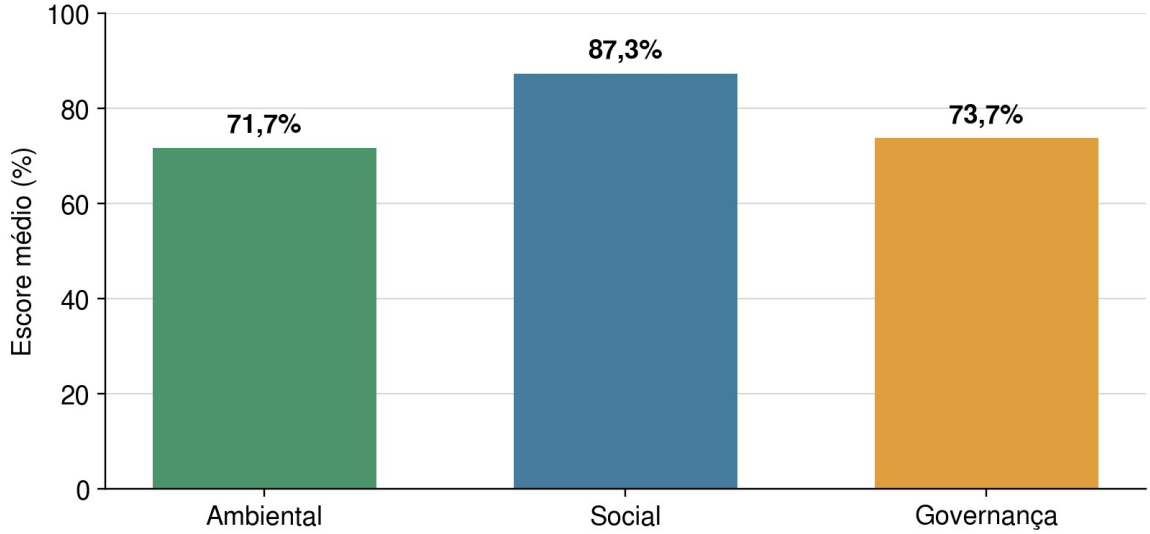
O escore global médio foi 77,6%, com mediana de 86,7%. O desvio-padrão amostral, originalmente apresentado na escala de 0 a 1 como 0,258, corresponde a 25,8 pontos percentuais. A amplitude foi de 90 pontos percentuais. A mediana acima da média indica concentração de resultados altos combinada com alguns escores muito baixos. Essa assimetria é compatível com diferenças de capacidade de reporte e com a desigualdade na disponibilidade dos documentos examinados.

A dimensão Social apresentou média de 87,3%, seguida por Governança, com 73,7%, e Ambiental, com 71,7%. A diferença não demonstra que o desempenho social efetivo seja superior. Ela indica que, dentro dos cinco critérios escolhidos e das fontes utilizadas, programas sociais, treinamento, saúde e segurança, direitos humanos e relacionamento com comunidade ou clientes apareceram com maior frequência documental.

O resultado também pode refletir a amplitude dos enunciados. O item social S5, por exemplo, era atendido por indicadores ou programas ligados à comunidade, aos clientes ou ao impacto social. Como bastava um desses caminhos, a média elevada pode decorrer da formulação abrangente. O mesmo vale para G2, que reuniu comitê de auditoria, sustentabilidade ou ESG. Essa característica aumenta a sensibilidade para localizar alguma estrutura, mas reduz a especificidade do diagnóstico.

Na dimensão Ambiental, a média foi reduzida principalmente pelo item de meta climática quantitativa com prazo. A existência de informações sobre energia, resíduos e água foi mais comum do que a combinação de meta, quantificação e horizonte temporal. Em Governança, as principais lacunas apareceram em remuneração variável vinculada a ESG e asseguração externa, práticas que exigem integração formal a incentivos e verificação independente.

Figura 1 - Escore médio por dimensão do checklist



Fonte: elaboração própria a partir da base da pesquisa.

4.3 Resultados por item

Tabela 4 - Resultado médio e distribuição por item

Item	Tema	Média	Sim	Parcial	Não/N/D
E1	Emissões de GEE	72,0%	27	5	9
E2	Meta climática quantitativa	42,7%	14	7	20
E3	Água	78,0%	30	4	7
E4	Resíduos/circularidade	80,5%	31	4	6
E5	Energia renovável/eficiência	85,4%	33	4	4
S1	Diversidade	81,7%	30	7	4
S2	Saúde e segurança	86,6%	31	9	1
S3	Treinamento	85,4%	34	2	5
S4	Direitos humanos/fornecedores	87,8%	36	0	5
S5	Comunidade/clientes/impacto	95,1%	38	2	1
G1	Conselho e independência	81,7%	31	5	5
G2	Comitê formal	92,7%	38	0	3
G3	Ética, anticorrupção e denúncia	87,8%	33	6	2
G4	Remuneração vinculada a ESG	58,5%	21	6	14
G5	Asseguração externa	47,6%	18	3	20

Nota: a última coluna reúne as duas categorias com pontuação zero, evitando interpretação indevida da distinção original.

Fonte: elaboração própria a partir da base da pesquisa.

Os maiores escores foram observados em S5, com 95,1%, e G2, com 92,7%. Em seguida aparecem S4 e G3, ambos com 87,8%, além de S2, com 86,6%. Esses

itens estão ligados à presença de programas, políticas, estruturas ou canais que costumam ser divulgados em relatórios e páginas institucionais. A interpretação deve considerar que alguns critérios aceitam alternativas amplas e, por isso, não medem o mesmo grau de profundidade.

Os menores resultados foram E2, com 42,7%, G5, com 47,6%, e G4, com 58,5%. Os três exigem evidência mais específica: meta climática quantitativa e datada, declaração de asseguração ou auditoria e vínculo explícito entre remuneração variável e metas ESG. A diferença sugere que a divulgação de políticas gerais está mais disseminada do que mecanismos formais de mensuração, verificação e responsabilização.

Do ponto de vista gerencial, esse padrão oferece uma agenda objetiva de aprimoramento documental. Empresas podem verificar se metas possuem linha de base, indicador e prazo; se relatórios são submetidos à asseguração independente; e se incentivos executivos incluem métricas ESG claramente descritas. O checklist é mais útil quando direciona esse aprofundamento do que quando reduz a análise a uma nota final.

Do ponto de vista metodológico, os itens de maior média precisam ser revistos quanto ao risco de facilidade excessiva. Uma versão futura deve separar comunidade, clientes e impacto social; distinguir comitê de auditoria de comitê de sustentabilidade; e decompor direitos humanos e gestão de fornecedores. Critérios mais específicos permitiriam identificar com precisão qual prática foi evidenciada e reduzir a heterogeneidade interna de cada pergunta.

4.4 Consistência interna e leitura de sensibilidade

O alfa de Cronbach foi 0,934 no conjunto dos 15 itens. Por dimensão, os valores foram 0,811 para Ambiental, 0,867 para Social e 0,793 para Governança. Em uma leitura convencional, os coeficientes indicariam consistência satisfatória ou elevada. Entretanto, o checklist é multidimensional e agrega temas que não são manifestações equivalentes de um único construto. Por isso, o resultado não deve ser apresentado como prova de validação.

A análise de sensibilidade reforça a cautela. A correlação entre o escore global e a cobertura documental original foi r = 0,956, e o alfa calculado apenas sobre a matriz binária de disponibilidade chegou a 0,893. Isso indica que parte importante da covariação entre itens pode refletir uma característica geral das empresas: possuir ou não documentos acessíveis e detalhados. Companhias com melhor infraestrutura de reporte tendem a receber respostas positivas em vários itens ao mesmo tempo.

Esse achado não torna o instrumento inútil. Pelo contrário, confirma que ele captura uma dimensão relevante de transparência e organização documental. O problema surge apenas quando essa dimensão é tratada como desempenho ESG integral. A denominação mais precisa é 'índice de aderência documental aos critérios do checklist'. A validade de conteúdo precisa ser examinada por especialistas e a confiabilidade da codificação, por concordância entre avaliadores.

Também é possível que itens amplos aumentem correlações porque a presença de um relatório robusto fornece evidência para várias perguntas. Uma empresa com relato integrado, formulário de referência completo e políticas públicas tende a pontuar em todos os blocos. Estudos futuros podem analisar correlações item-total, redundância, estrutura dimensional e estabilidade em nova amostra, mantendo a interpretação formativa do índice.

Tabela 5 - Consistência interna e verificação de cobertura

Escala	Itens	Alfa	Leitura recomendada
Checklist completo	15	0,934	Exploratório; possível efeito de cobertura
Ambiental	5	0,811	Consistência satisfatória
Social	5	0,867	Consistência elevada
Governança	5	0,793	Consistência satisfatória
Matriz binária de cobertura	15	0,893	Disponibilidade documental também covaria

Nota: os coeficientes não demonstram validade de conteúdo, validade externa ou concordância entre avaliadores.

Fonte: elaboração própria a partir da base da pesquisa.

4.5 Implicações e limites da aplicação-piloto

A aplicação confirma que é possível converter referências extensas em um roteiro curto de busca documental. As 615 classificações foram registradas em estrutura única, permitindo identificar lacunas e comparar dimensões. Essa operacionalidade é a principal evidência produzida pelo estudo. Ela sustenta a continuidade do desenvolvimento, mas não a adoção imediata do instrumento como rating.

A primeira limitação é o protocolo de fontes. A base reuniu documentos de anos diferentes e, em alguns casos, apenas os anexos disponíveis no momento da análise. Uma nova rodada deve realizar busca padronizada diretamente nas fontes corporativas e oficiais. Sem esse procedimento, diferenças entre empresas podem refletir a quantidade de material recuperado, não apenas a transparência organizacional.

A segunda limitação é a regra de classificação. 'Não' e 'N/D' receberam a mesma pontuação, mas seus significados são distintos e nem sempre foram aplicados de modo uniforme. Para corrigir o problema, cada item deve possuir uma árvore de decisão. Se uma fonte afirmar explicitamente que a prática não existe, utiliza-se 'Não'. Se a evidência não for localizada, utiliza-se 'N/D'. Se o item for materialmente inaplicável, utiliza-se 'N/A' com justificativa e exclusão do denominador.

A terceira limitação é a ausência de validade de conteúdo e concordância interavaliadores. O instrumento foi derivado de referenciais reconhecidos, mas isso não substitui julgamento formal de especialistas. Recomenda-se submeter cada item a painel de profissionais e pesquisadores, calcular Índice de Validade de Conteúdo e realizar dupla codificação de uma parcela dos documentos. Divergências devem ser resolvidas por protocolo previamente definido.

A quarta limitação é a materialidade. O mesmo conjunto de 15 itens foi aplicado a cinco setores e recebeu pesos iguais. O resultado permite apenas uma leitura genérica de evidenciação. Versões setoriais devem incluir temas específicos e pesos justificados. Até que essa calibração ocorra, a comparação setorial ou empresarial deve ser apresentada como descritiva e provisória.

Por fim, a rastreabilidade precisa ser ampliada. O registro ideal inclui nome completo do documento, ano, URL, data de acesso, página, trecho que sustenta a decisão e limite organizacional da evidência. Esse conjunto possibilita auditoria,

replicação e atualização do diagnóstico. A transparência do próprio checklist deve ser compatível com a transparência que ele pretende avaliar.

5. Conclusão e contribuições

O estudo desenvolveu e examinou a aplicação-piloto de um checklist ESG documental com 15 itens, distribuídos igualmente entre Ambiental, Social e Governança. O instrumento foi aplicado a 41 companhias listadas na B3 e gerou 615 classificações. A análise confirmou sua operacionalidade para organizar evidências públicas, produzir um retrato agregado e indicar temas que exigem aprofundamento.

A média global foi 77,6%, com mediana de 86,7% e desvio-padrão de 25,8 pontos percentuais. A dimensão Social apresentou maior média, seguida por Governança e Ambiental. Os resultados mais baixos concentraram-se em meta climática quantitativa, asseguração externa e remuneração vinculada a metas ESG. Esses achados descrevem a evidenciação observada nas fontes da aplicação e não autorizam conclusões sobre efetividade socioambiental ou superioridade de empresas e setores.

A contribuição teórica está em reforçar a separação entre divulgação, aderência documental e desempenho ESG. A contribuição metodológica consiste em apresentar um instrumento rastreável e uma leitura crítica de seus próprios limites. O alfa elevado demonstrou associação entre os itens, mas a sensibilidade à disponibilidade documental mostrou que consistência interna não equivale à validação.

A principal conclusão é que o checklist possui potencial como ferramenta de triagem e preparação para avaliações mais completas. Para uso gerencial, pode orientar a organização de documentos, a identificação de lacunas e a elaboração de planos de ação. Para uso acadêmico ou comparativo, ainda requer protocolo uniforme de busca, critérios específicos por resposta, rastreabilidade completa, validação por especialistas, concordância entre avaliadores e calibração de materialidade setorial.

Pesquisas futuras devem reaplicar o instrumento em ano-base único, com duas avaliações independentes, e comparar os resultados a indicadores externos, como relatórios assegurados, ratings ou métricas operacionais. Também se recomenda separar itens amplos e testar versões por setor e porte empresarial. Com essas etapas, o checklist poderá evoluir de aplicação-piloto para instrumento validado de diagnóstico documental.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR ISO 26000:2010: diretrizes sobre responsabilidade social. Rio de Janeiro: ABNT, 2010.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT PR 2030-1:2024: ambiental, social e governança (ESG): conceitos, diretrizes e modelo de avaliação e direcionamento para organizações. Rio de Janeiro: ABNT, 2024.

BERG, F.; KÖLBEL, J. F.; RIGOBON, R. Aggregate confusion: the divergence of ESG ratings. Review of Finance, v. 26, n. 6, p. 1315-1344, 2022. DOI: 10.1093/rof/rfac033.

CDP. CDP full corporate questionnaire overview: 2024. London: CDP, 2024. Disponível em: https://cdn.cdp.net/cdp-production/cms/guidance_docs/pdfs/000/005/005/original/CDP-full-corporate-questionnaire-overview_-_2024.pdf. Acesso em: 12 jun. 2026.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, v. 16, n. 3, p. 297-334, 1951. DOI: 10.1007/BF02310555.

ELKINGTON, J. Cannibals with forks: the Triple Bottom Line of 21st century business. Oxford: Capstone, 1997.

GLOBAL REPORTING INITIATIVE. GRI 1: fundamentos 2021. Amsterdam: GRI, 2021. Disponível em: https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-portuguese-translations/. Acesso em: 12 jun. 2026.

GLOBAL REPORTING INITIATIVE. GRI 3: temas materiais 2021. Amsterdam: GRI, 2021. Disponível em: https://www.globalreporting.org/publications/documents/english/gri-3-material-topics-2021/. Acesso em: 12 jun. 2026.

INSTITUTO BRASILEIRO DE GOVERNANÇA CORPORATIVA. Código das melhores práticas de governança corporativa. 6. ed. São Paulo: IBGC, 2023.

LONDON STOCK EXCHANGE GROUP. Environmental, social and governance scores from LSEG. London: LSEG, 2024. Disponível em: https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf. Acesso em: 12 jun. 2026.

NUNNALLY, J. C.; BERNSTEIN, I. H. Psychometric theory. 3. ed. New York: McGraw-Hill, 1994.

OECD. OECD Business and Finance Outlook 2020: sustainable and resilient finance. Paris: OECD Publishing, 2020. DOI: 10.1787/eb61fd29-en.

TAVAKOL, M.; DENNICK, R. Making sense of Cronbach's alpha. International Journal of Medical Education, v. 2, p. 53-55, 2011. DOI: 10.5116/ijme.4dfb.8dfd.

UNITED NATIONS. Guiding principles on business and human rights: implementing the United Nations 'Protect, Respect and Remedy' framework. New York; Geneva: United Nations, 2011.