Accountability quantified

What 26 years of GAO reports can teach us about government management
About the authors


**Steve Berman** is a member of Deloitte Consulting LLP's Advanced Analytics and Modeling group. He has more than 24 years of experience, including over 16 in consulting, in many industries, including insurance, retail, consumer and industrial products, life sciences, automotive, and telecommunications. Berman typically assists companies by using analytics to gain insights into the business and by designing models that improve corporate efficiency or pricing.

**Vishwa Kolla** is a manager in Deloitte's Advanced Analytics and Modeling practice. Kolla's industry experience includes work in retail, technology, media, telecommunications, and insurance. He focuses on using modeling and analytical techniques to understand the customer life cycle in its entirety—acquisition, behavior, revenue and profitability drivers, and retention. Kolla has an MBA from Carnegie Mellon University and an MS in computer science.

**William D. Eggers** is a director with Deloitte Services LP, where he leads Deloitte's public sector research. He is the author of eight books, including his latest (with Paul Macmillan), *The Solution Revolution: How Business, Government, and Social Enterprises are Teaming Up to Solve Society's Toughest Problems* (Harvard Business Review Press, 2013). Follow him on Twitter at @wdeegers or contact him by email at weggers@deloitte.com.

Acknowledgements

The technical expertise of **Jon Vanden Bosch** and **Corey Long** was absolutely crucial to the successful completion of this project. Additionally, this project could not have been completed without the support of Deloitte's Advanced Analytics and Modeling Center under **Mike Petrillo, James Nason**, and **Justin West**. We also thank **Gabriel Sudduth** for his input on the report.

Illustration by **Santiago Uceda**
Contents

Introduction | 2

Question 1: Are GAO recommendations an effective way to drive targeted change within agencies? | 4

Question 2: What problem areas are less likely to see successful implementation of recommendations? | 6

Question 3: Where do agencies most often succeed in implementing GAO recommendations? | 7

Question 4: Do agencies improve their success in implementing recommendations if they receive repeated recommendations in the same area? | 9

Question 5: Have major events or wider government trends substantially altered GAO’s core focus over time? | 11

Question 6: Did the Government Performance and Results Act (GPRA) result in recommendations that were difficult to implement? | 13

Question 7: Have outside events like the Obama administration’s management goals or terrorism had a targeted impact on GAO’s focus? | 14

Practical steps: Treating oversight recommendations as data | 17

Appendix: Methodology | 19

Endnotes | 21
HOW can a government leader transform tens of thousands of pages of oversight reports from the US Government Accountability Office (GAO), their agency’s own Inspector General, and any number of other sources into a clear action plan for his or her agency? What problems have defied past intervention? Are there parts of the organization that are succeeding where others are failing?

These are crucial questions for every government leader to answer, but doing so can seem impossible, given the sheer volume of oversight material. Fortunately, text analytics creates the opportunity to understand what core themes emerge from voluminous reports. This study aims to assist government leaders both by drawing management insights from 1.3 million pages of GAO reports and by using GAO as an example of how agencies can better structure their internal oversight activities to quantify accountability and drive results.

We analyzed the 40,000-plus recommendations made by GAO to federal agencies from 1983 through 2014. Each of these recommendations is made on the basis of detailed, evidence-based GAO reports meant to spur specific actions within agencies. These recommendations span almost every issue in government, from information security to environmental resources. Using text analytics, we asked seven key questions to assess GAO’s effectiveness as a change agent in government and to understand what areas GAO tends to focus on. Our findings are summarized in the sidebar titled “Summary of questions and answers” (see the appendix for a detailed account of our methodology).

GAO’s recommendations are important, not only because of their content, but also because GAO’s categorization and reporting structures offer many lessons for government leaders seeking to set up systems to quantify accountability. In “whack a mole oversight,” agencies confront whichever problem seems to be most salient at the moment without a precise view into what issues have consistently plagued multiple departments or which problems have been the most resistant to remediation. Compounding this difficulty is that the information used to fix the identified problem too often consists of anecdotes driven by the personal experiences of the people tasked with solving the problem.

The antidote to this is to quantify the effectiveness of accountability by adopting a structured process to treat oversight reports as text-based data. Agencies can accomplish this transition by digitizing and categorizing their recommendations, auditing whether recommendations were followed, text-mining their recommendations for insights, and sharing results of implementations in real time. Successfully accomplishing this transition lifts the “shroud of darkness” that surrounds quantifying the efficacy of oversight mechanisms and helps agency heads drill down into problem areas. As such, our findings have implications for GAO, Congress, the executive branch, and government agencies everywhere.

What is GAO?

The Government Accountability Office (GAO) is an independent, nonpartisan agency that works for Congress. Often called the “congressional watchdog,” GAO investigates how the federal government spends taxpayer dollars. GAO’s goal is to help improve the performance and ensure the accountability of federal agencies. Each year, it regularly publishes reports, testimony, and issue recommendations which identify problems within the US government.
What is text analytics?

Text analytics is the process of obtaining meaningful insights from large compendiums of documents using a computer. Before text analytics, massive amounts of text could only be partially understood by individual people painstakingly reading small portions of the available information. Now, text analytics allows for insights to be drawn from everything from millions of streaming tweets to mountains of complex documents. Text analytics is not a cure-all. When reading a small number of documents, it can be less precise than a human reader and usually cannot understand how complex concepts relate to one another as well as a human could. However, text analytics is the only way to process extremely large bodies of text and, when combined with domain expertise, can deliver dramatic results. Text analytics is responsible for major breakthroughs ranging from IBM’s Watson to helping the US government fight terrorism.

SUMMARY OF KEY QUESTIONS AND ANSWERS

Question 1. Are GAO recommendations an effective way to drive targeted change within agencies?
Yes. GAO recommendations have an 81 percent success rate.

Question 2. What problem areas are less likely to see successful implementation of recommendations?
Mentions of the “data,” “high-ranking officials,” and the presence of cross-agency issues were all associated with lower success rates than other comparable recommendations.

Question 3. Where do agencies most often succeed in implementing GAO recommendations?
Recommendations pertaining to information security, information technology, education, and equal opportunity tended to have the highest success rates.

Question 4. Do agencies improve their success in implementing recommendations if they receive repeated recommendations in the same area?
No. Repeated examination in a particular area does not improve implementation rates.

Question 5. Have major events or outside groups substantially altered GAO’s core focus over time?
No. Five to seven core topics have consistently dominated GAO’s recommendations over time.

Question 6. Did the Government Performance and Results Act (GPRA) result in recommendations that were difficult to implement?
No. In fact, GPRA-related recommendations have a 2 percent higher success rate than all other non-GPRA related recommendations.

Question 7: Have outside events like the Obama administration’s management goals or terrorism had a targeted impact on GAO’s focus?
There was a limited impact. The greatest influence was seen in an increased focus on transparency. Outside events such as the attacks of September 11, 2001 have driven changes in GAO content.
YES. Overall, during the period between 1983 and 2008, 81 percent of GAO’s recommendations were successfully completed by federal agencies.¹ This figure is based on GAO’s own audit of the outcomes of its recommendations to agencies, which is described in detail in the sidebar “How GAO scores implementation of its recommendations.”

The high success rate of implementing GAO recommendations has been consistent over time. The “worst” year was in 1992, when federal agencies recorded a 76 percent completion rate, just 5 percent off their overall average of 81 percent. The best year was 2001, when federal agencies recorded an 86 percent completion rate (figure 1).

Unfortunately, it can sometimes take agencies more than four years to implement a GAO recommendation. GAO could address this issue by setting target completion dates for implementing each recommendation and then making real-time data available to the public showing how long it is taking each agency to implement GAO recommendations. This could motivate agencies to more quickly address GAO recommendations and realize the benefits they deliver to the public. In addition to setting specific deadlines, GAO could further motivate agencies by classifying both noncompliance and extreme tardiness as failures. If this is not done, it is easy to lose a sense of urgency, and recommendations can languish. Also, GAO could consider giving recommendations a “criticality” score that allows its analysts to sort out whether an agency is struggling with major items or just “nice-to-have” items. The lack of such a score for assessing the importance of an individual recommendation is

---

**Figure 1. Completion rates for GAO recommendations over time**

Note: Size of ticker indicates the total number of recommendations GAO made in each year.
a weakness in GAO’s methodology, and the agency would likely benefit from adopting one.

**Bottom line:** Agencies consistently implement GAO recommendations. Given GAO’s finite resources, it cannot uncover and solve every problem—but it is successful at driving its recommendations to fruition within agencies. Having said that, the low variability in the implementation rate could indicate that GAO may, for better or for worse, sometimes be issuing recommendations that it believes agencies are likely to implement.

---

### How GAO scores implementation of its recommendations

When GAO conducts an audit, it issues reports which often make specific recommendations to an agency based on evidence found during the audit. Completion of these recommendations is tracked by the GAO staff that made it. Before GAO considers a recommendation complete, a GAO staffer must obtain documentation from the agency demonstrating change, such as conducting interviews with key agency stakeholders, or otherwise obtain proof that the recommendation was implemented. When an agency completes a recommendation, the GAO staff mark it as complete. If it lapses for more than four years or the agency refuses to comply, it is marked as not completed in the system. This high degree of due diligence gives us confidence that these recommendations are genuinely being acted on.

### What GAO’s record-keeping process teaches us about how to quantify oversight effectiveness

GAO has been able to quantify agency compliance by diligently following up year after year with agencies to see if recommendations have come to fruition. GAO’s consistent and persistent approach underpinned our ability to conduct the multi-decade analysis above. Agencies which hope to quantify the success of their own internal oversight initiatives will need to commit to a similar level of effort.

Tracking the success of recommendations is not only crucial for gaining an overall view of the oversight’s efficacy but also improves the insights text analytics can produce. For example, if an agency clearly defines success for their recommendations, they would not only know how the contents of their recommendations had changed over time but would understand how the contents of their *successful* recommendations had changed.

Recommendations should set specific deadlines and classify both noncompliance and extreme tardiness as failure. Without this, it is easy to lose a sense of urgency, and recommendations can languish. Also, consider giving recommendations a “criticality” score that allows analysts to sort out whether the agency is struggling with major items or just “nice to haves.”
Question 2: What problem areas are less likely to see successful implementation of recommendations?

AGENCIES that are given a GAO report/recommendation with any of the following attributes are less likely to succeed in implementing the recommendation than they would be for other comparable recommendations:

- Data are part of the problem
- Multiple agencies are required to coordinate on an issue touching multiple topics
- High-ranking officials or Congress (especially the Appropriations Committees) are mentioned
- Health care or transportation are part of the problem

Overall, these findings reinforce well-known trends in the government. Problems that cut across multiple agencies, impact numerous parts of an agency's duties, or deal with high-profile issues can be difficult to address, even with the help of a detailed GAO report outlining the core issues that need to be addressed. Such problems are often grounded in many of the known root issues that cut across the federal government. While GAO may be helpful in diagnosing them, outside bodies (such as Congress, the Office of Management and Budget, and the White House) may have to engage if deep-seated issues with problematic characteristics are to be resolved.

Bottom line: Whenever a recommendation cuts across multiple topic areas or multiple agencies, agency leaders should devote additional financial and political resources to it if they reasonably expect the recommendation to be implemented. To improve the chances of implementation, GAO could prioritize laying out more-specific roadmaps for issues relating to data, multiple agencies, or the involvement of higher-ranking officials. Policymakers should also not be surprised if such issues persist in the absence of larger structural reforms.

What GAO’s record-keeping process teaches us about how to quantify oversight effectiveness

GAO has gone to significant lengths to provide its reports and recommendations in a format that is easy for text analytics programs to read. This structure made it possible to extract the presence of “data issues,” “Congress,” and “high ranking officials” in the analysis above. Agencies that want to be able to track mentions of particular important issues need to invest in creating a well-structured electronic database of their reports. This investment will be a one-time cost that underpins all future text analytics efforts.
Question 3: Where do agencies most often succeed in implementing GAO recommendations?

AGENCIES have the highest success rate in implementing GAO recommendations in four key areas: information security, education, information technology, and equal opportunity.

Information security comes in first, with a near-perfect completion rate of 94 percent. Given the frequent and high-profile information security and information technology failures in the US federal government, it is important to characterize the recommendations in this category. Rather than calling for large system implementation changes, GAO’s recommendations related to information security are often tactical. For example, the Securities and Exchange Commission was encouraged to “adequately back up critical data files on key workstations used for storing large accounting data files and ensure that...

What the recommendation success rate teaches us about how to quantify oversight effectiveness

Managers of federal agencies need to be alert when success rates get too high or too low. Low completion rates may indicate an inability or an unwillingness to comply with oversight. Conversely, if completion rates are too high, it could be that the problem under examination has been solved and no longer needs investigating, or that the demands being made do not address the elephant in the room. A stable and high completion rate in the presence of known ongoing problems is a red flag that recommendations may not be addressing the real issues.

Figure 2. Top four most “successful” areas for implementing GAO recommendations

<table>
<thead>
<tr>
<th>Topic</th>
<th>Completion rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information security</td>
<td>94 percent</td>
</tr>
<tr>
<td>Education</td>
<td>87 percent</td>
</tr>
<tr>
<td>Information technology</td>
<td>87 percent</td>
</tr>
<tr>
<td>Equal opportunity</td>
<td>87 percent</td>
</tr>
</tbody>
</table>

Figure 3. Suggested actions based on issue criticality and recommendation completion rate

<table>
<thead>
<tr>
<th>Low success rate in implementing recommendations</th>
<th>High success rate in implementing recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes Invest energy in understanding and fixing the root causes of repeated failures.</td>
<td>Push the envelope to make sure no stone is left unturned in a crucial area.</td>
</tr>
<tr>
<td>No Determine if oversight body is asking too much of a low-risk area and adjust emphasis appropriately.</td>
<td>Study success to understand what can be used to improve completion rates in problematic areas.</td>
</tr>
</tbody>
</table>
mission-critical application contingency plans contain key information." This does not mean that GAO never issues large-scale directives, but it does point to the fact that its recommendations are generally within the reach of the agencies it evaluates. Considering the government's overall track record in information security, this is an area where GAO might be able to be more aggressive in the scope of its future recommendations.

Equal opportunity and education were also areas of success. With respect to equal opportunity, this success is likely partially due to the fact that public fallout coming from a failure to comply with a discrimination/equal opportunity issue can be significant. Education's high success rate may be primarily because many education recommendations made requests for simpler actions such as updating information or dealing with grantees where an agency had natural leverage.

**Bottom line:** GAO's high success rate in the information technology space indicates it may have room to increase the number and strength of the specific recommendations it gives around IT security issues. The federal government as a whole should take pride that, when identified, equal opportunity issues appear to often be addressed in a timely manner.
Question 4: Do agencies improve their success in implementing recommendations if they receive repeated recommendations in the same area?

Over time, GAO has categorized its recommendations into areas ranging from very broad (national defense) to the extremely specific (“Oil importation within the Department of Energy”). Ideally, if GAO were to give agencies repeated recommendations in small and specific areas like IT acquisition, we would see agencies’ rate of successfully implementing these recommendations rise as they address the root cause of the problem. In actuality, that is not the case. There is no meaningful relationship between how many recommendations an agency receives in a specific area and how often they succeed in that area. In other words, an agency seems no more likely to implement a recommendation in the “information systems” category whether it receives 100 or 500 recommendations in that category.

The chart below shows the relationship, or rather the lack thereof, between the number of recommendations given to a particular agency in a specific area and the agency’s success rate in implementing those recommendations. To avoid including areas within agencies that can be swayed by just a few recommendations, only agencies with at least 50 recommendations in a targeted area are included. Similarly, agencies that received more than 1,000 recommendations in a given area (like DoD’s thousands of recommendations relating to military forces) are excluded because the recommendations are too broad to be easily interpreted. The finding of “no relationship” was consistent almost regardless of where we placed these cutoffs.

We further investigated whether GAO increasing the number of its recommendations issued to an agency led that agency to improve its success rate over a longer period of time. Specifically, we compared the number and success rate of recommendations from the 1990s to the 2000s. If an area (like IT acquisition in the Department of Energy) received increased recommendations in the 2000s compared to the 1990s, one would assume that the success rate in that area would also increase. This was not the case. Instead, there was no statistically significant relationship between an agency receiving an increased number of recommendations in a specific area and an improved completion rate.

The notion that some problem areas, no matter how many GAO recommendations target them, persist in being a challenge is supported by GAO’s high-risk list. GAO’s high-risk list tracks core areas that pose material issues to the US government. Of the 30 areas currently on GAO’s high-risk list, 16 have been on the list for more than a decade despite receiving consistent attention. Out of a total of 55 high-risk areas ever identified, only 23 (less than half) have been resolved. These 23
Success rate of implementation

Figure 4. Repeated recommendations to an agency in the same area do not improve an agency’s success rate

Source: Deloitte Research data analysis.

resolutions represent real successes, but the remaining issues point to the difficulty GAO has in compelling agencies to resolve deeply entrenched issues.

**Bottom line:** Over time, we would expect that agencies would learn from past experience and show higher completion rates in areas where they consistently receive recommendations to improve. However, this does not appear to be the case, at least in the quarter century’s worth of data that we examined. Given agencies’ generally high success rate in resolving individual GAO recommendations, Congress should view GAO as an effective scalpel but not a panacea for the federal government’s longstanding problems. GAO may sometimes succeed in helping agencies make meaningful changes, but problems often exist that are beyond GAO’s reach. Addressing the root causes of these problems (with the recent issues in veterans’ health care at the Veterans’ Administration as a prime example) may often require Congressional intervention, as well as a sustained focus on changing the culture within an organization.

**What targeting specific areas for change teaches us about how to quantify oversight effectiveness**

A well-formed system of categorizing reports (also known as a taxonomy) not only provides decision makers quick insight into where recommendations are succeeding but also how many recommendations are being made in key areas. Maintaining a consistent taxonomy over time allows leaders to quantify both the focus and long-term impact of their interventions.
The simple answer is—not really. All of the top four most common categories for recommendations from the 1980s, 1990s, and 2000s were exactly the same: audit, government operations, law enforcement, and national defense. Coming in fifth in the 1980s and 1990s was the environment. In the 2000s, the creation of the Department of Homeland Security, an increased focus on international affairs, and the rise of information technology supplanted natural resources and the environment. However, the environment ranked eighth in the 2000s, indicating that its consistent staying power remained somewhat intact. Overall, this stability is particularly notable because these top five categories make up over 51 percent of all GAO recommendations.

As a whole, our results point to how consistently these areas have been top of mind for GAO from 1983 through 2009. This does not

Figure 5. Top five areas of recommendations in the 1980s, 1990s, and 2000s

Top five areas of recommendations, 2000–2009

<table>
<thead>
<tr>
<th>Category</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>National defense</td>
<td>2,851</td>
</tr>
<tr>
<td>Government operations</td>
<td>2,638</td>
</tr>
<tr>
<td>Auditing and financial management</td>
<td>2,305</td>
</tr>
<tr>
<td>Justice and law enforcement</td>
<td>1,505</td>
</tr>
<tr>
<td>Information technology</td>
<td>1,383</td>
</tr>
</tbody>
</table>

Top five areas of recommendations, 1990–1999

<table>
<thead>
<tr>
<th>Category</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>National defense</td>
<td>2,097</td>
</tr>
<tr>
<td>Auditing and financial management</td>
<td>1,408</td>
</tr>
<tr>
<td>Justice and law enforcement</td>
<td>1,373</td>
</tr>
<tr>
<td>Government operations</td>
<td>1,320</td>
</tr>
<tr>
<td>Natural resources and environment</td>
<td>850</td>
</tr>
</tbody>
</table>

Top five areas of recommendations, 1980–1989

<table>
<thead>
<tr>
<th>Category</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justice and law enforcement</td>
<td>1,284</td>
</tr>
<tr>
<td>Government operations</td>
<td>1,213</td>
</tr>
<tr>
<td>Auditing and financial management</td>
<td>1,045</td>
</tr>
<tr>
<td>National defense</td>
<td>1,013</td>
</tr>
<tr>
<td>Natural resources and environment</td>
<td>557</td>
</tr>
</tbody>
</table>

What 26 years of GAO reports and text analytics can teach us about government management
What the GAO’s consistency of focus teaches us about how to quantify oversight effectiveness

Oversight managers need to balance steadfastness with “blowing in the wind.” A successful oversight program should have some constancy, as we see in the consistent emphasis of GAO. However, if there are not some shifts, it is likely a sign that oversight has become unresponsive to the changing landscape of problems presented to the agency or government as a whole.

Beyond measuring shifts in the number of times recommendations are made in a given category, oversight managers can use text analytics to detect whether the composition of recommendations in a category is changing. If the terms being discussed within a category change drastically, management should consider altering their taxonomy. This will prevent their staff from force fitting recommendations into categories which have fundamentally changed in meaning.

Bottom line: The top four issue areas investigated by GAO stayed constant throughout three decades. These top-line trends transcended party control and presidential management agendas.
Question 6: Did the Government Performance and Results Act (GPRA) result in recommendations that were difficult to implement?

GPRA was passed in 1993 with the intent to improve government performance management. More recently, the GPRA Modernization Act of 2010 established a new framework and new processes aimed at encouraging a more crosscutting and integrated approach to focusing on results and government performance management. In order to comply with GPRA, agencies must create strategic plans and set performance goals, among other things, and increase their overall focus on improving agency management. Because its intent is to create effective governance structures, GPRA stands out as a natural area where, given the breadth of recommendations pertaining to it, completion rates among the 2,020 recommendations related to the act might be lower than average. But this was not the case. The success rate of GPRA-related recommendations was 83 percent—two percentage points higher than the overall average.

Part of the reason for this success rate may be that GPRA-related recommendations were not meaningfully more likely to stem from broad multi-agency problems than other recommendations. Each GPRA recommendation, on average, touched 3.1 agency components versus the overall average of 2.8. 

**Bottom line:** GPRA-related recommendations have historically been no more difficult to implement than other recommendations.

What the success rate of implementing GPRA recommendations teaches us about how to quantify oversight effectiveness

Oversight leaders can use their database of recommendations to test whether assumptions about the difficulty of certain initiatives are actually true. This will help dispel myths about “impossible tasks” and identify where true difficulty resides.

Text analytics can also help add context to why these unexpected successes or failures are occurring. For example, an agency head could investigate whether problems surrounding “metrics” are described differently in recommendations that succeed than in recommendations that fail.
IN the area of transparency, yes. On other topics, no. President Obama’s FY2010 budget outlined the administration’s management priorities in six key areas:

1. Replacing the Program Assessment Rating Tool (PART)
3. Transforming the federal workforce
4. Managing across sectors
5. Reforming federal contracting
6. Transparency, technology, and participatory democracy

For each of these areas, we analyzed whether the administration’s initiatives altered the focus of GAO reports. Even though GAO formally works for Congress, not the administration, we would expect that the interests of the president would influence the interest of some members of Congress. In five out of the six areas, we saw no meaningful increase in the percentage of reports that mentioned terms associated with the administration’s focus. (Stimulus-related terms did increase during the Obama administration, but these mentions were deemed insignificant because of the small number of reports related to economic stimulus in the preceding years; moreover, it would have been impossible for GAO to investigate a program that did not exist during previous administrations.) Other priorities, such as federal contracting and the federal workforce, have been a perennial focus of GAO regardless of a particular administration’s focus. It is thus difficult to elevate these priorities above their historic position of interest. Transparency was the lone area that saw a marked increase in GAO recommendations during the Obama administration (figure 6).

Beyond the administration’s formally stated goals, we investigated whether the prevalence...
of any key terms in GAO’s summaries had declined during the Obama administration. A material decline would indicate that an area had received less emphasis during the Obama administration than during the Bush administration.

In general, a great deal stayed the same from the Bush administration to the Obama administration, with more than 60 percent of all commonly occurring terms fluctuating in frequency by less than 20 percent. This indicates that most common keywords did not change in frequency. The top nine terms that did decline in frequency (“decliners”) are shown in figure 7.

The decline in many of these terms indicates a reduced emphasis on terrorism and accounting. The decline of “terrorism” as a key term in GAO recommendations has been gradual since the term’s peak in 2003. The fact that the use of the term “terrorist” peaked in 2003 is logical, because it would have taken GAO time

<table>
<thead>
<tr>
<th>Term</th>
<th>Percent decline from Bush to Obama administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditor</td>
<td>77%</td>
</tr>
<tr>
<td>Corporation</td>
<td>63%</td>
</tr>
<tr>
<td>Generally accepted</td>
<td>59%</td>
</tr>
<tr>
<td>Accounting</td>
<td>38%</td>
</tr>
<tr>
<td>Productivity</td>
<td>37%</td>
</tr>
<tr>
<td>Attack</td>
<td>36%</td>
</tr>
<tr>
<td>Computer</td>
<td>34%</td>
</tr>
<tr>
<td>Fire</td>
<td>34%</td>
</tr>
<tr>
<td>Terrorist</td>
<td>32%</td>
</tr>
</tbody>
</table>

In general, a great deal stayed the same from the Bush administration to the Obama administration, with more than 60 percent of all commonly occurring terms fluctuating in frequency by less than 20 percent. This indicates that most common keywords did not change in frequency. The top nine terms that did decline in frequency (“decliners”) are shown in figure 7.

The decline in many of these terms indicates a reduced emphasis on terrorism and accounting. The decline of “terrorism” as a key term in GAO recommendations has been gradual since the term’s peak in 2003. The fact that the use of the term “terrorist” peaked in 2003 is logical, because it would have taken GAO time

<table>
<thead>
<tr>
<th>Term</th>
<th>Percent decline from Bush to Obama administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditor</td>
<td>77%</td>
</tr>
<tr>
<td>Corporation</td>
<td>63%</td>
</tr>
<tr>
<td>Generally accepted</td>
<td>59%</td>
</tr>
<tr>
<td>Accounting</td>
<td>38%</td>
</tr>
<tr>
<td>Productivity</td>
<td>37%</td>
</tr>
<tr>
<td>Attack</td>
<td>36%</td>
</tr>
<tr>
<td>Computer</td>
<td>34%</td>
</tr>
<tr>
<td>Fire</td>
<td>34%</td>
</tr>
<tr>
<td>Terrorist</td>
<td>32%</td>
</tr>
</tbody>
</table>
to write terrorism-focused recommendations for agencies after the attacks of September 11, 2001.

The decline in the discussion of accounting predates the Bush administration and appears to be part of a general trend at GAO toward examining government management issues and enforcing “accountability” instead of just enforcing accounting standards. Additionally, because the government has received over time more and more clean audit opinions, the underlying need for GAO accounting reports may have diminished.

**Bottom line:** It is possible for the focus of the executive branch, as well as outside events, to alter what GAO investigates. However, we should not expect dramatic changes in the overall composition of GAO reports. Instead, highly specific and differentiated components of the administration’s priorities (such as increasing transparency) or major current events appear more likely to change what GAO investigates.

---

**What this teaches us about how to quantify oversight effectiveness**

Broad categorizations of reports are crucial to understanding general areas of emphasis and success in oversight actions. For example, understanding if recommendations related to “information technology” are failing at a high rate and have been steadily increasing in number, would indicate this area is ripe for root cause analysis. Text analytics can support this analysis by unearthing trends that fall outside of predetermined categories. Continuing our example, it could be that all recommendations mentioning “human resources” are failing at a higher rate across the entire organization but that the IT department has borne the brunt of it. This means that, while IT may be the most pronounced case, the root cause of the problem (HR) is actually damaging the entire organizations but just to a lesser extent. Without text analytics, it might not have become clear this issue was cross-cutting and the scope of the intervention could have been too narrow.
As demonstrated by the above analysis, text analytics is a powerful tool for extracting insights from the massive amounts of copy contained in GAO oversight reports. To capitalize on analytics to gain a more nuanced understanding of oversight effectiveness, GAO and other federal agencies should begin to look at GAO and Inspector General reports as data that can be programmatically analyzed and made more transparent. (The same idea applies for state and local governments.) While GAO has taken many steps in this direction, other agencies can follow suit at their own internal oversight organizations by considering the following steps:

- **Keep score.** Agencies should look to mimic GAO by investing in “keeping score”—that is, tracking where their recommendations are succeeding or failing. They should track both the importance of a recommendation as well as the timeliness of implementation in order to drive results in the most significant places. Following GAO’s example, agencies should assign a point of contact to each recommendation. This small but helpful practice greatly aids those seeking supplemental details for particularly relevant recommendations.

- **Convert reports to a text analytics-friendly electronic format.** Agencies should plan on a one-time investment in converting reports into an electronic format that is friendly to text analytics. This will allow for analyses that go beyond filtering reports along the lines of preset categorizations and enable an agency to respond to specific leadership interest in how oversight has been changing over time. An electronic archive of reports will also allow for some historical analysis of oversight even if the documents have not been coded into predefined categories.

- **Establish a coding structure for reports.** To facilitate text analytics, agencies should establish a coding structure for reports that has a stable core but retains the ability to adapt to changes in management interest and the influence of current events.

- **Uncover hidden trends.** Agencies can use text analytics to look for trends that predefined categories may fail to reflect or reveal.

- **Develop a standard taxonomy for oversight reporting terms.** Agencies wishing to create a single view of all oversight initiatives will need to create cross walks between GAO’s categorizations and any other sources of oversight they wish to combine. This will allow data to be easily “translated” between different categorization systems to allow for a consolidated view. We recommend that GAO work with other agencies to develop a central standard “base” taxonomy that other agencies can use as a starting point for their own custom taxonomies.

- **Develop real-time accountability scorecards—and make them public.** To drive speed of implementation and enhance accountability, agencies can make real-time dashboards that score recommendation completion rates and speed of
implementation accessible to the appropriate audiences. If the information in the dashboards is not too sensitive, agencies should share these dashboards with the public and Congress so that successes and difficulties can be honestly discussed.

To illustrate a straightforward example of a public-facing accountability dashboard, we’ve created an interactive dashboard that displays agencies’ success rate in completing GAO recommendations over time. The dashboard allows users to select almost any combination of agencies, topics, and/or time periods of interest. Additionally, the actual text of some GAO recommendations is displayed at the bottom of the dashboard. (Please note that these data were pulled in 2014 and are static, so there will be some discrepancies between the information on GAO’s website and that contained in this dashboard). The dashboard can be found in the online version of this article at http://dupress.com/articles/text-analytics-and-gao-reports.

In total, the interactive dashboard contains thousands of unique charts. It can be embedded as an object in blogs, so please feel free to use it as a starting point for your virtual conversations. Let us know what you think with the hashtag #AccountabilityQuantified or by tweeting @du_press.

If you are a journalist or academic researcher who has an interest in performing analyses of GAO reports, drop us a line at publicsectorresearch@deloitte.com and we can work with you on questions that you are interested in. These data, and the analytical work that we’ve done to make it usable, are open assets that we want to continue to build on with other researchers.
To begin the analytical process, “success” had to be clearly defined. Fortunately, GAO codes each recommendation it makes as “Open,” “Closed—not implemented,” and “Closed—implemented.” For our review, we defined a “successful” recommendation as one that is coded “Closed—implemented,” and a failed recommendation as one whose status is “Closed—not implemented” or “Open” (provided that the report was more than five years old). Recommendations whose status was unclear (8 percent of the sample) or less than five years old were not included in our averages. Also, if the same recommendation was given to two different agencies, we counted it as two separate recommendations. This population represented less than 7 percent of our sample. It should be noted that some GAO reports made no recommendations and so were not included in deriving our recommendation-focused findings, while other reports made multiple recommendations. In the latter case, all of the report’s recommendations were included in our analysis. However, prior to the 2000s, much of the text was not easily machine-readable. Therefore, summaries of reports, which were provided by GAO in text format, were used instead. This caused a substantial reduction in the 1.3 million pages of full text that could have been analyzed, as optical character recognition was not perfectly reliable.

The full text of GAO’s reports does not suffer from the four-year-plus time lag that recommendations do. Consequently, all available full reports, regardless of their age, were included in the report-focused portions of our analysis.

To help summarize the text into a form that could be used in a mathematical model, the text of each report title, recommendation, and report body in the sample was extracted and analyzed. This process allowed a number of variables to be created based on concepts contained in the text. These text variables indicate the presence or absence of a specific keyword or phrase in each report title, report body, or recommendation. Terms and phrases that appeared too infrequently or too frequently to meaningfully analyze were dropped from the analysis.

In addition to the text variables drawn from the text of the reports, the study also considered a number of other factors related to each report, as well as certain conditions existing at the time of the report, as potential explanatory variables. Examples of these other variables include the specific agencies related to a report, variables indicating political control of legislative or executive branches at report issuance, and characteristics of the report itself, such as its age and length. In total, these efforts yielded 213 variables that were included in our analysis.

While examining the recommendation data as individual recommendations has its merits, examining them solely on this basis ignores a critical component of their context: The same agencies receive multiple recommendations over time, providing a consistent way to segment the data every year. This makes the annual recommendation rates dependent on one another from a time series standpoint. To treat them as such, we collapsed the data into a time series and then tested each variable for statistical significance using a panel time
This technique resulted in the elimination of a small number of observations, but still left us a large number of observations spanning from 1983 to 2008.

After beginning the process with 213 independent variables, we were left with 54 significant variables. For the sake of readability, we will not list all of the significant variables here. Instead, we have grouped them into general findings in the write-up above. These variables successfully explained 55 percent of the year-over-year change in agencies’ rates of complying with GAO recommendations. As such, the findings discussed here do not explain all possible reasons that a recommendation may or may not have been taken, but they do provide enough statistical power to tell us a great deal.

We invite researchers interested in more details on our work to reach out to publicsectorresearch@deloitte.com.
Endnotes

1. As described in the appendix, this research defined "successful" recommendations as those that GAO scored as “Closed—implemented,” and a “failed” recommendation as those that GAO scored “Closed—not implemented” or “Open” (provided that the recommendation was more than five years old).

2. The careful reader will note that “data” was a key driver of difficulties in question 2, but that IT issues often showed success. This is possible because not all “data”-related issues stem from IT difficulties; they may stem from process issues or other sources.


4. As mentioned before, this study truncates the analysis of recommendation implementations at 2008 due to the significant time it takes agencies to respond to recommendations. Because of this, any changes in these patterns pertaining to PART since the start of the Obama administration could not be identified.


6. This was excluded from the analysis because it did not contain sufficient detail to generate a concrete list of related keywords.

7. Agencies with fewer than 200 recommendations were excluded, as a small number of recommendations can produce success rates that are artificially high or low due to a small but overly influential set of recommendations.

8. Only topics at the top of GAO’s taxonomy were selected. Parties interested in sub-categories should reach out to the authors directly.

9. For this study, the text mining software Megaputer was used to parse text, identify keywords, and create indicators.

10. The data were averaged by agency and by year. As such, each variable became the “percent” of recommendations that fell into a given agency’s batch of recommendations per each year. A Hausman test was performed to differentiate between fixed and random effects modeling, and fixed effects were chosen at better than the .01 p- value level. An F test was performed to test for the appropriateness of variable reduction, and the data passed at the .1, .05, and .01 significance levels. Examination of residual plots from the final models confirmed an absence of outliers and confirmed model integrity. A two-way fixed effects model was estimated to account for significant variation at both the agency and year level.

11. A logistic regression was fit to the original un-aggregated data, but failed the Hosmer-Lemeshow test for lack of fit and so could not be accepted.

12. Significance was at the p < .05 level for each individual independent variable.
Contacts

Daniel Byler
Data scientist, Deloitte Services LP
+1 571 814 7166
dbyler@deloitte.com

John Houston
Principal, Deloitte Consulting LLP
National practice leader
Advanced Analytics and Predictive Modeling
+1 617 437 3993
jhouston@deloitte.com

John Lucker
Principal, Deloitte Consulting LLP
Market leader
Global Advanced Analytics and Modeling
+1 860 725 3022
jlucker@deloitte.com
About Deloitte University Press

Deloitte University Press publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte University Press is an imprint of Deloitte Development LLC.

About this publication

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte LLP and its subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.

Copyright © 2015 Deloitte Development LLC. All rights reserved.

Member of Deloitte Touche Tohmatsu Limited