

Innovating the IPCC review process—the potential of young talent

Lianne van der Veer · Hans Visser · Arthur Petersen · Peter Janssen

Received: 5 September 2013 / Accepted: 4 May 2014 / Published online: 20 May 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract There is significant potential in young talent for enhancing the credibility of the scientific assessments such as the IPCC's by contributing to quality assurance and quality control. In this essay, we reflect on an experiment that was done by the Dutch government as part of its government review of a contribution to the IPCC's Fifth Assessment Report (AR5). In an effort to review the entire Working Group II contribution to the AR5 within the official review period for the Second Order Draft (SOD), the PBL Netherlands Environmental Assessment Agency turned to PhD students. This article shows that a systematic review focusing on transparency and errors of a large scientific assessment document using young talented scientists can be successful if certain conditions are met. The reviewers need to have intrinsic motivation to conduct the review. There needs to be a communication plan that fosters engagement and a clear methodology to guide the reviewers through their task. Based on this experiment in review, we reflect on the wider potential for openness and crowdsourcing in scientific assessment processes such as the IPCC's.

1 Introduction

In January 2010, the media reported two errors in specific parts of the Working Group II (WGII) contribution to the Fourth Assessment Report of 2007 by the Intergovernmental Panel on Climate Change (IPCC). The media reports gave rise to questions about the credibility of the overall IPCC assessment. The commotion led to a passionate debate in the Dutch parliament on the credibility of the IPCC. In this debate a motion that focused on regaining the credibility of the IPCC found broad support in parliament (The Dutch House of Representatives 2010). The PBL Netherlands Environmental Assessment Agency (hereafter "PBL")¹ was asked to conduct a review of the regional chapters of the WGII report published

¹The PBL Netherlands Environmental Assessment Agency is an independent governmental body that by statute provides the Dutch Government and Parliament—and the European Commission, European Parliament and UN organizations—with scientific advice on problems regarding the environment, sustainability and spatial planning. PBL authors have contributed to several assessments of the IPCC. During the Third and Fourth Assessments, the PBL hosted the Co-Chair and Technical Support Unit of Working Group III (Mitigation of Climate Change).

L. van der Veer · H. Visser (✉) · A. C. Petersen · P. Janssen
PBL Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands
e-mail: Hans.Visser@pbl.nl

A. C. Petersen
Institute for Environmental Studies (IVM), VU University, Amsterdam, The Netherlands

in 2007 as well. In its resulting evaluation report (PBL 2010), the PBL offered several recommendations for producing credible assessment reports, not only within the IPCC but more widely across a range of scientific assessment activities for policy making.

One of the recommendations in PBL (2010) for scientific assessment processes such as IPCC assessments was to investigate the possibilities of pre-publication “crowdsourcing.” Since finding errors, notably in numbers, is a time-consuming effort for authors, reviewers, and management, who already are limited in terms of time and resources, it was suggested that the possibility of using “crowdsourcing” technologies as part of the review during the last stages of completion could be considered. Crowdsourcing has the advantage of providing a large reservoir of voluntary resources to be tapped. A disadvantage identified by the PBL was the amount of resources needed to manage and moderate tasks that would be handed out to the public.

Another recommendation in PBL (2010), pertaining in particular to intergovernmental scientific assessment processes such as the IPCC’s was that governments, which have an important responsibility in the review process, should seriously invest in the government/expert review of drafts. It was suggested by the PBL that apart from the reviewing of issues of specific national interest, governments, assisted by independent experts, could review provenance and foundations for statements in the draft versions of the

Summaries for Policymakers. The PBL (2010) evaluation provides an example of how this could be done methodologically.

In this essay, we reflect on an experiment that was done by the Dutch government as part of its government review in which we gave follow-up to a combination of these two recommendations, that is, we employed crowdsourcing in a much intensified government review. Based on this experiment in review, we reflect in particular on the wider potential for openness and crowdsourcing in scientific assessment processes such as the IPCC’s.

What will become clear in the course of this essay is that there is significant potential in young talent for enhancing the credibility of the IPCC by contributing to quality assurance and quality control. For its Fifth Assessment Report (AR5), the IPCC had already decided to support the more senior lead authors and reduce work burden for them, while simultaneously ensuring that all numerical statements in chapters were scrutinized more closely than ever, by inviting young volunteer “chapter scientists” to assist the authors in this task. This internal review and fact checking can be seen as an important step in ensuring traceability to the cited literature. In addition to this enhanced internal review process, the PBL decided to also conduct a review of IPCC’s WGII contribution to the AR5 that was systematically organized to trace issues of transparency and consistency and made use of young talent.

IPCC Assessment Reports undergo three review rounds where the second round is the one where besides individual experts also governments may comment on the Second Order Draft (SOD) of the full report, its Summary for Policymakers (SPM) and its Technical Summary (TS). Eight weeks were available for this review of the WGII AR5 report, from 29 March to 24 May 2013. Reviewing the SOD of the WGII report left the PBL with the imposing task to review a report of two summaries, 30 chapters, totaling 2,701 pages within the official review period of 8 weeks. After subtracting time for organization and preparing the review for submission a short 4 weeks period remained to do the actual review. Faced with this challenge PBL searched for an innovative solution, building on the two recommendations that it had given earlier.

By connecting both recommendations the PBL realized that it did not necessarily need people with vast experience to conduct a review that focuses on transparency, consistency and

errors. All that was needed were people with affinity to the subject and high analytical skills. Therefore, the team responsible for the Dutch government review decided to crowdsource a component of the review. Given current limitations with respect to confidentiality of IPCC draft, a delimited group of young scientists—PhD students—was targeted in a structured crowdsourcing process.

In designing the review experiment two parallel tasks were envisaged. The first task was to guarantee a systematic and high quality review. The second task was to engage the reviewers. All methods needed for performing the first task were designed prior to the review. The second task had a two-stage approach. First, the reviewers were recruited. Second, a communications and engagement plan was designed to suit the pool of reviewers we composed. Therefore, we start this essay by describing the recruitment process. This is followed by a description of the methods used to ensure engagement of the reviewers and a high quality review. In the subsequent section we give the results of the review experiment, in process terms. We make use of the insight gained in how the reviewers experienced the review from a survey that was distributed 2 weeks after the review deadline (69 out of 90 reviewers filled in this survey). Subsequently, we reflect on the wider potential for openness and crowdsourcing in scientific assessment processes such as the IPCC's.

2 Recruiting reviewers for the PBL experiment

Before recruiting the PhD-student reviewers for the PBL government review experiment a profile was composed for the “ideal reviewer.” The reviewers had to be intrinsically motivated to review the IPCC report, have excellent analytical skills and pursue a doctorate in a field related to the areas covered by the IPCC assessments.

To ensure intrinsic motivation it was important to recruit reviewers who shared the review mission to improve the quality of IPCC reports. Reviewers who share the review mission “pursue [the review] goals because they perceive intrinsic benefits from doing so” (Besley and Ghatak 2003). A shared mission ensures intrinsic motivation that makes it possible to recruit talents against a much lower cost (Gailmard 2010). Hence, a recruiting letter was used that emphasized the mission and goals of the review.

To make sure the reviewers had excellent analytical skills and were pursuing a doctorate in a relevant field the first batch of PhD students were recruited via the Research School for Socio-Economic and Natural Sciences of the Environment (SENSE), a joint national graduate school of the environmental sciences departments of 11 universities in the Netherlands. At present 600 PhD students are members of SENSE. The goal was to find two students per chapter, thus 60 in total, who were willing to invest two or 3 days for doing a review. Within a few days after the recruiting e-mail was sent to all SENSE members, 25 people applied. But after that, applications stopped.

To guarantee a good review an additional number of 35 PhD students was needed. Therefore, the search was made public by using twitter. Via @MachteldAnna the tweet was plugged in a social media network on climate change. Within 2 weeks there were over 120 applications from people who were willing to join in the review. After checking whether the applicants were pursuing a doctorate in a relevant field, the reviewers were selected based on their educational background and work

experience. The final team consisted of 90 PhD students, 40 taking their PhD in The Netherlands and 50 taking their PhD abroad. The group represented 40 nationalities from 6 continents, see Table 1.

3 Methods applied in the PBL experiment

3.1 Fostering engagement

One of the main issues with this review was to make sure the reviewers were motivated throughout the review period. To accomplish this, insights from communication theory were used to design an engagement plan. The plan builds on two pillars: communication and incentives.

3.1.1 Communication

Because over half of the reviewers live abroad, e-mail was the main means of communication. To ensure that e-mail communication with the reviewers would help engage them, a communication plan was written based on knowledge and experience from the field of customer support. Research in that field shows that reliability, responsiveness and competence are the main variables in establishing a durable relationship. Reliability includes accuracy, correctness and consistency of communication. Responsiveness consists of both general response time and the speed in which problems are addressed. Competence focuses on the knowledge and skill the service provider shows in answering questions and solving problems in a satisfactory manner (Parasuraman et al. 1985; Yang et al. 2004).

Table 1 Country of origin for the 90 reviewers

Asia	27	Africa	16	Europe	33
Bangladesh	3	Cameroon	1	Belgium	1
China	5	Ethiopia	4	Cyprus	1
India	8	Kenya	2	Denmark	1
Indonesia	1	Malawi	1	France	1
Iran	4	Nigeria	3	Great Britain	2
Malaysia	1	South-Africa	1	Italia	5
Nepal	1	Trinidad	2	Netherlands	17
Pakistan	1	Uganda	1	Poland	1
Taiwan	1	Zimbabwe	1	Portugal	1
Vietnam	2			Spain	1
				Switzerland	1
				Turkey	1
North America	7	Oceania	3	South America	4
Canada	1	Australia	2	Bolivia	1
Mexico	1	New Zealand	1	Brazil	1
United States	5			Colombia	1
				Peru	1

Response time is an important variable because people who use e-mail for communication, expect a quick response. If e-mails are not answered within the 48 h time window, customers satisfaction with the service provided and the engagement with the organization will reduce significantly (Mattila and Mount 2003; Strauss and Hill 2001). To keep the reviewers engaged all e-mails had to be answered within a 48-hour window.

To guarantee reliability standardized e-mails that contained both the information needed at certain moment and a preview of what was about to happen in the near future were used. Before sending the e-mails they were altered to suit the specific characteristic of the recipient, as personalization shows interest in the e-mailer and increases the experience of being taken seriously (Strauss and Hill 2001).

Competence was assured by letting a communication expert handle all e-mails, backed up by two experts with long experience in reviewing IPCC reports in case that the communication expert was unable to provide an answer.

3.1.2 Incentives and motivation

A prerequisite for the success is that all reviewers are motivated, and stay motivated during the process, to do the job assigned to them, within the time limits given, while using the prescribed methodology. The first step to ensure motivation was recruiting intrinsically motivated people. Besides that, a test and a number of incentives were used to make sure the reviewers would be motivated until the end of the review period.

To check for motivation we tested the responsiveness of our reviewers. This meant that reviewers received an email requesting information on their educational history (bachelor/master, and PhD institute). If they did not respond to this e-mail they had to be removed from the review team. To stimulate that the full review period was finished successfully, additional incentives were introduced. Students affiliated with SENSE gained educational credits if they finished the review successfully. A kickoff meeting and a seminar were organized to provide network opportunities, interesting speakers, feedback and free food. To introduce a competitive element the top 10 reviews got the chance to present their findings to an expert panel² in a seminar. This panel awarded the best review with 250 euro per person. To persuade the reviewers to read the chapter text thoroughly and look for mistakes a “blooper bonus” of 50 euro per person was introduced for the team who found the factual mistake with potentially the most media impact. To inform reviewers living abroad who could not attend the planned meetings, videos were made available to them of all speakers at the kick-off meeting and the seminar, providing moreover a live stream-recording of the seminar.

3.2 Realizing quality

Motivated reviewers *alone* do not make a quality review. PBL set out a path to guide the students in their review. Furthermore, PBL performed a quality check to filter all review comments as a final step.

3.2.1 Guiding the reviewers

After selecting the 90 reviewers, three had to be assigned to every chapter of the report. These three students received the Summary for Policymakers (SPM) and Technical Summary (TS)

² The expert panel consisted of Leo Meyer (PBL and Head, Technical Support Unit for the IPCC Synthesis Report), Maarten van Aalst (Lead Author for the WGII AR5 report), Hans Visser and Peter Janssen (both senior scientists at PBL).

and the chapter they were assigned to review via e-mail. The reviewers did not receive the full IPCC report, because this report was still confidential. They had to cooperate and were asked to send their comments combined in one Excel file. The choice of a specific chapter was based on the reviewers' preferences, their educational background or work experience and their nationality. Their nationality was used to distribute the regional chapters, as local knowledge also contributes to their expertise too.

A complicating factor to the distribution was that every team had to contain at least one Netherlands-based reviewer. This guaranteed that every team could delegate at least one person to the meetings. It was expected that the drop-out risk of Dutch students would be smaller due to better incentives. Spreading the Netherlands-based reviewers thereby improved the chance that all chapters would be reviewed.

To achieve a good quality review PBL provided a methodology to the reviewers. This methodology was derived from the review PBL carried out on the regional chapters of the IPCC Working Group II contribution to the Fourth Assessment Report (PBL 2010; Strengers et al. 2013). Based on experiences of this review four aspects appeared important: (i) scientific comprehensiveness, depth and balance, that is including the full range of views, (ii) transparency and consistency of messages throughout the report, (iii) readability and (iv) factual mistakes.

The first aspect, scientific comprehensiveness, depth and balance, is best addressed by specialized scientists. We organized this expert review parallel to the PhD students review (not further discussed in this essay). The main focus of the PhD students was directed to the second aspect, the transparency and consistency of the messages throughout the report. They were instructed to trace back all statements in the SPM and TS which pointed to their chapter, down to the executive summary (ExSum) of their chapter, to the main text of the chapter and, finally,

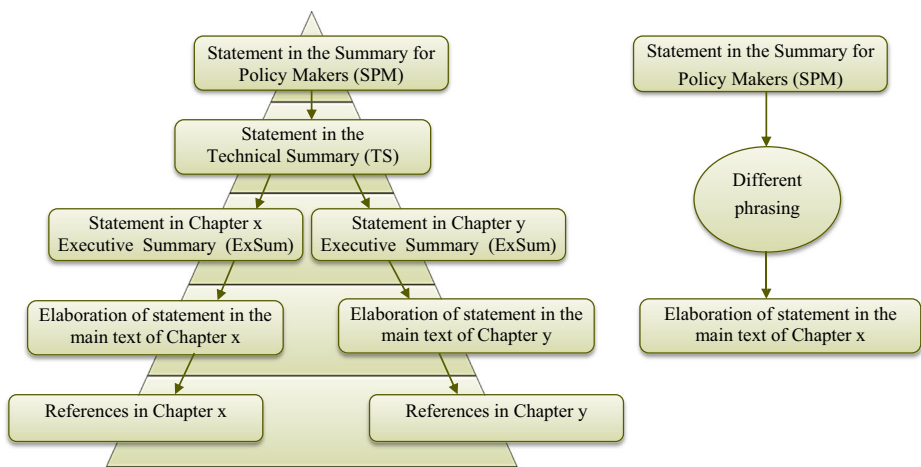


Fig. 1 Five levels of information can be distinguished in the IPCC WGII report. The pyramid at the left shows the ideal situation: a statement at the highest level (SPM) is founded in the Technical summary, the Executive Summary of Chapter x, and, possibly, Chapter y. Further down, the statement is founded in the main text of the Chapter, and at the lowest level, in the references. The diagram on the right illustrates a situation which is not ideal: a statement in the SPM is founded in the main text of Chapter x, with no references to peer-reviewed literature. Furthermore, the statement in the SPM has been rephrased. In the consistency check any statement applying to the specific chapter under review is followed up and down in the pyramid. However, care should be taken if a statement in the SPM and TS refers to statements in more than one Chapter (the 'Chapter x' and 'Chapter y' paths in the pyramid)

to check as many literature references as time permitted. Because the reviewers only received one chapter checking cross-references between chapters was not possible for them. See Fig. 1 for an illustration. Besides this check for transparency and consistency, reviewers were asked to read the summaries and their chapter text in order to comment on readability and to find factual mistakes.

Next to this screening approach the reviewers were asked to categorize all their comments into nine error/comment codes: E1a (Errors that can be corrected by a straightforward replacement of text), E1b (Errors that require a redoing of the assessment of the issue at hand), E2 (Inaccurate referencing), C1 (Insufficiently substantiated attribution), C2 (Insufficiently founded generalization), C3 (Insufficiently transparent expert judgment), C4 (Inconsistency of messages), C5 (Untraceable reference) and C6 (Unnecessary reliance on grey referencing). Further details of this categorization are given in PBL (2010, pp. 29–32) and Strengers et al. (2013) and are not repeated here.

To instruct the reviewers, both a written review-methodology guide and a live explanation of the review methodology during the kick off meeting were provided. A video of the live explanation was made available so that reviewers who were unable to attend the meeting could watch it.

We note here that the categorization of comments using codes such as given above, is not part of the normal IPCC governmental review methodology. However, it could help IPCC authors to pre-screen the tenths of thousands of comments they receive after a review period. E.g., one can split a review-Excel-spreadsheet directly into a file with small errors (E1a), a file with large errors (E1b), or a file with comments concerning insufficiently founded generalizations (C2). Furthermore, we do not claim that comments that follow this categorization are more important for the governmental review than comments which focus on scientific comprehensiveness, depth and balance, that is including the full range of views. Instead we claim that comments focusing on our categorization also need to be dealt with in the review and we acknowledge that possibly other avenues than the formal governmental review could be used to deliver such comments.

3.2.2 *Quality control*

To ensure that all student comments were acceptable for the Dutch government review two senior scientists of PBL were asked to conduct a full review of all comments. The error/comment codes provided to the reviewers were aimed to facilitate this quality check. All comments were grouped into three categories: (i) suitable for the Dutch government review, (ii) minor comments (e.g. typos) and (iii) not suitable for submission. All the comments in the first category were checked for correctness and edited where necessary before including them in the government review. Comments in the second category were sent to the IPCC in an informal way without a further quality check. Comments of the third type were discarded. Some of these latter comments had to do with the fact that the students had to follow only their own chapter and had no means to check statements in other chapters (cf. left panel in Fig. 1).

4 Results of the PBL experiment

4.1 Fostering engagement

4.1.1 *Communication*

The communication plan we made at the start of the review was followed quite closely. During the 4 weeks reviewers sent around 500 e-mails to PBL. 90 % of these e-mails were replied to

within a 48-hour time window and all were personalized. The fast response time was reached by answering e-mails in weekends and holidays too. To provide the team with information, PBL sent bulk e-mails addressed to all reviewers. These e-mails were not personalized.

The survey results showed that 78 % of the reviewers who asked a question via e-mail were very satisfied with the responses; 5 % was not satisfied. The communication within the review process was not a 100 % success, however. Halfway through the review period, every team was asked how they were doing. All teams responded to the e-mail and three teams reported internal struggles. From the three teams whose internal struggles prevented them from finalizing a review of acceptable quality, two teams did not report their problems when asked.

4.1.2 Incentives and motivation

All incentives were applied as planned. A proof of participation was added to the incentives because reviewers not affiliated with the SENSE research school requested it for their curriculum vitae.

An indication of the motivation and responsiveness of the reviewers can be obtained from the response of reviewers when they were asked to return an e-mail with the confidentiality clause (as required by the IPCC procedures) within a few days. Over 85 out of 90 reviewers returned the confidentiality clause within 4 days. The others followed after one reminder. This illustrates that all reviewers were responsive to e-mails and willing to participate.

The videos posted online were hardly watched. The only video that had a significant amount of views was the video explaining the review methodology. The live stream was watched by almost half the people who were not present at the final meeting. But the reviewers did not rate it very high. Based on the feedback of the reviewers it can be concluded that the video and sound quality of the live stream was insufficient. A drawback from using internet to watch videos and live stream is that some reviewers were unable to watch them, due to internet restrictions in their country (e.g., blocked access).

We started the review process with 90 students; five reviewers did not make it to the end. Two dropped out due to personal circumstances and three were unresponsive. Over half the reviewers invested more time in the review than we had asked them to do. Of the 69 reviewers that filled in the questionnaire, 61 replied that they would like to be part of the team that will review the Synthesis Report in 2014. The eight reviewers who did not want to participate in the next review were all reviewers based in the Netherlands. The educational credits and the meetings were only available for the reviewers in the Netherlands. The survey results showed that time invested by the Netherlands-based reviewers was similar to the time invested by reviewers based abroad. The dropout rate was similar for both groups.

These additional incentives did not seem to influence the investment made by the reviewers. Whether this supports the idea that intrinsic motivation is the main driver for the reviewers to join the review-experiment remains unclear since we did not interview them on their intrinsic motivation to join in. Neither did we explicitly check what role the educational credits played for the SENSE students as incentive to participate in the review. It is striking, though, that all eight reviewers that did not want to join the new review – which would not provide extra credits – were SENSE students.

In addition to the incentives that were explicitly built into the process, it is worth mentioning that participating in such an experiment can have additional benefits for young scientist in their early career, which could add to their motivation: it gives them an opportunity

to be part of a key scientific activity, offering an interesting experience in collaboration and learning from others in conducting successful reviews.

4.2 Realizing quality

4.2.1 Guiding the reviewers

Ideally every chapter should be reviewed by someone with basic knowledge of the academic literature on the subject. In reality the 90 reviewers did not form a well-balanced group. For instance, there were ten hydrologists and no reviewers with a specialty in the polar region. In the end, most students did receive a chapter in their area of expertise or a regional chapter corresponding to their native country. In Section 3.2 we described four issues that had to be addressed in a complete review. The review results showed that especially the second issue, consistent messages throughout the report, was thoroughly addressed in the scope of the review. There were many comments on readability too, and the teams found at least five real factual errors with the potential to get media attention. As the reviewers only received their own chapter, and not the entire WGII report, this review did not comment on issues that cross over *between* chapters.

The expert panel² which judged the top 10 contributions at the seminar, was very surprised by the high quality of both reviews and presentations (the best ten teams were selected beforehand). The review teams showed that young talented PhD students are well able to check for transparency and consistency and in several cases also scientific comprehensiveness, depth and balance. Besides these ten best reviews, 15 other chapters were reviewed adequately. For the five remaining chapters the review was not sufficiently done.

It is hard to pinpoint which variables make the difference between a sufficient or an insufficient review. It was evident that the best review teams distinguished themselves by showing excellent analytical skills. Three of the teams that delivered an insufficient review had problems working together or had unresponsive teammates.

4.2.2 Final quality control

The review resulted in a total of 3,155 comments received. From this total, 1,407 comments (45 %) were included in the government review, 1,326 comments (42 %) were categorized as minor comments and 422 comments (13 %) were categorized as not suitable for submission to the IPCC. The upper panel of Fig. 2 shows these numbers for each of the 30 chapters separately. The nature of comments not suitable for submission was varying. Most of them were caused by reviewers who misunderstood, had not read the underlying literature or ignored that when a particular SPM and TS statement points to multiple chapters (which were not made available to them), the underpinning will not necessarily be fully found in the chapter assigned to them (cf. left panel of Fig. 1).

It can be concluded that more comments does not guarantee better quality. For instance, the reviews for Chapter 7 (Food security) and Chapter 22 (Africa) were considered best by the expert panel at the seminar. Both these chapters had less than average amounts of comments. Chapter 19 (Emergent risks and key vulnerabilities) and 26 (North America) had the highest number of comments but those reviews were considered weak.

4.3 Dutch government reviews for AR4 and AR5 compared

It is interesting to compare the present AR5 review with the Dutch government's review of the SOD of the IPCC's WGII contribution to the Fourth Assessment Report (AR4) that was

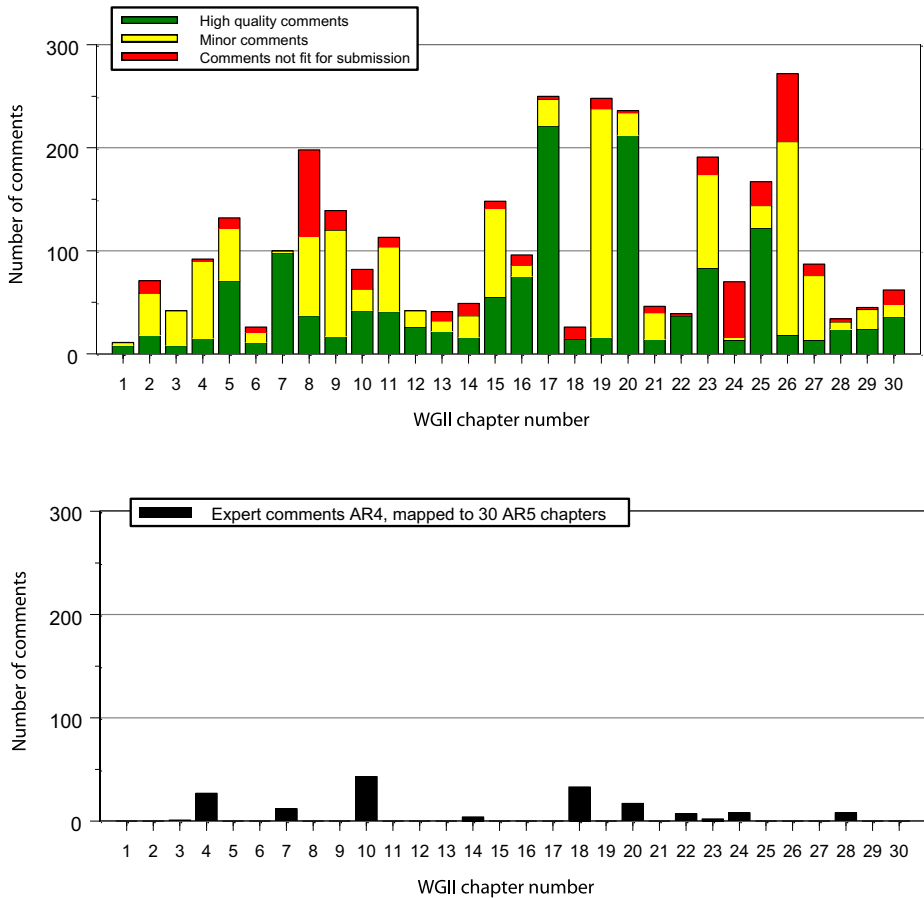


Fig. 2 Number of review comments per WGII chapter (upper panel). Comments were grouped into three categories: high quality comments (green, 1,407 in total), minor comments (yellow, 1,326 in total) and comments not fit for submission to IPCC (red, 422 in total). The lower panel shows the number of review comments per WGII chapter, given by Dutch experts in 2006 for the Second Order Draft of the WGII report. We have reallocated comments for the 20 chapters in 2006 to the 30 chapters in 2013 to visualize differences between AR4 and AR5 comments. The lower panel has 162 comments in total

submitted in 2006. The AR4 review was made by experts who were free to comment on any chapter they wanted. No specific guidelines were given to them. The comparison first shows that the total number of Dutch government comments submitted has increased by nearly an order of magnitude: from 162 in 2006 to 1557 in 2013. It also becomes clear that in 2006 only half the chapters were commented on in the Dutch government review.

For illustrative purposes we have mapped the comments of Dutch experts from 2006 (20 chapters) onto the most closely related chapter (out of 30 chapters) of the AR5. The result is given in the lower panel of Fig. 2. The graph shows that (i) the number of comments is much lower than those shown in the upper panel of Fig. 2, and (ii) many chapters were not commented on at all. The latter was a consequence of not having coupled experts to the full range of chapters in the Dutch government review in the previous IPCC assessment round.

We categorized the AR4 and AR5 comments as to the four aspects named in Section 3.2: (i) comprehensiveness, depth and balance, (ii) transparency and consistency, (iii) readability and

(iv) factual mistakes. The percentages are 35, 9, 48 and 8 %, respectively. The distribution for the AR5 PhD student comments are 20, 18, 35 and 28 %, respectively. The latter percentages are estimated from the 1,407 student comments submitted to the IPCC (the green bars in the upper panel of Fig. 2). For the PhD student comments in the AR5 experiment the number of comments on consistency and transparency nearly equal the number of comments on comprehensiveness, depth and balance, both around 20 %. In the 2006 expert comments, as we expected, much less attention was paid to consistency and transparency (only 9 % of the comments, versus 35 % for comments on comprehensiveness, depth and balance).

5 Discussion and conclusion

What the results of the PBL review experiment show is that one can fruitfully employ young talent through crowdsourcing in checking and ensuring the transparency and consistency of scientific assessment reports. Our specific approach based on two parallel tasks seems to have been effective for its purpose: both the approach for realizing quality and the approach to foster engagement can be deemed successful. Through fostering engagement we insured a minimum dropout of five reviewers. The time the reviewers invested in the review was often more than we had asked for. An additional advantage of our approach is that a group of 60 reviewers were willing to conduct a new review with PBL the following year.

A negative side of the approach chosen is the large investment in time to insure that all e-mails were answered within 48 h. This investment did not prohibit us from being surprised by teams who did not deliver good reviews due to team problems and unresponsive team members. And using internet for videos or live streaming is not ideal as two reviewers in countries with internet restrictions could not use these services.

Our approach for realizing quality resulted in a good-quality review for 25 out of 30 chapters. For the 10 best reviewed chapters the results were above expectations. Both the goal of a thorough check for consistency and finding errors in the style of “the Himalayan glaciers melting by 2035” were met. The resulting review was positively received by the IPCC as well. The leadership of WGII was pleased with the review and the Technical Support Unit agreed to mention all 90 reviewers as expert reviewers in the published report.

Next to these positive results we found that the review instructions, including the categorization of different error/comment codes, could be made more clear and more simple. Also the possibility for cross-checking with other chapters could perhaps be offered next time.

A point of discussion in utilizing young talent for increasing the credibility of the IPCC is whether there are also ways to cover the major quality issue of comprehensiveness, depth and balance. For instance, key scientific issues may have been omitted or not weighted heavily enough by a particular IPCC chapter. Here the limitation is that most of the young talent that may become available through even targeted crowdsourcing experiments will not be seasoned experts in the field. Therefore the approach demonstrated in this essay is not sufficient to fully assure and control quality in the assessment chapters.

We conclude that in the context of a government review of IPCC reports, when there are limited time and limited resources for an extensive review, PhD students are a valuable resource. Many of them are intrinsically motivated and their academic qualities make them excellent reviewers, but they need to be guided with care. More generally, we claim that young talent can be much more utilized in review processes than is presently the case, provided that besides paying attention to the methods for realizing quality especially there should be a well-prepared plan for fostering engagement.

This lesson is also relevant for the present discussions on the future of the IPCC. Some governments, among which the Netherlands, think that the IPCC needs to become more transparent in the future and provide more focused and up-to-date assessments. In terms of technology, this could be facilitated by making IPCC assessments fully web-based. Such digitalization would help to further increase the transparency of the reports. For example, more links could be included to the relevant parts of scientific publications to simplify the accessibility to the sources. But obviously making such changes should involve stepping up quality assurance and quality control, for which one would not want to overburden the authors. Hence, new innovative ways, possibly involving crowdsourcing young talent, should be explored to be able to provide more dynamic and more transparent scientific assessments.

Acknowledgments We thank the 90 PhD Students (for a list of names and countries, see: <http://www.pbl.nl/en/news/newsitems/2013/www.pbl.nl/sites/default/files/cms/PBL-2013-Names-and-affiliations-of-Contributors-Dutch-govt-review-SOD.pdf>), Leo Meyer, Hayo Haanstra, Maarten van Aalst and Laurens Bouwer for their contribution to the PBL review.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Besley T, Ghatak M (2003) Competition and incentives with motivated agents. LSE STICERD Research Paper No. TE465
- Gailmard S (2010) Politics, principal–agent problems, and public service motivation. *Int Public Manag J* 13(1): 35–45. doi:10.1080/10967490903547225
- Mattila AS, Mount DJ (2003) The impact of selected customer characteristics and response time on e-complaint satisfaction and return intent. *Int J Hosp Manag* 22(2):135–145. doi:10.1016/S0278-4319(03)00014-8
- Parasuraman A, Zeithaml VA, Berry LL (1985) A conceptual model of service quality and its implications for future research. *J Mark* 20:41–50
- PBL (2010) Assessing an IPCC assessment. An analysis of statements on projected regional impacts in the 2007 report. Netherlands Environmental Assessment Agency (PBL), The Hague/Bilthoven. PBL publication number: 500216002. To be downloaded from: <http://www.pbl.nl/sites/default/files/cms/publicaties/500216002.pdf>
- Strauss J, Hill DJ (2001) Consumer complaints by e-mail: an exploratory investigation of corporate responses and customer reactions. *J Interact Mark* 15(1):63–73. doi:10.1002/1520-6653(200124)15:1 <63
- Strengers BJ, Meyer LA, Petersen AC, Hajer MA, van Vuuren DP, Janssen PHM (2013) Opening up scientific assessments for policy: the importance of transparency in expert judgements. PBL working paper 14. PBL Netherlands environmental assessment agency, The Hague. To be downloaded from: <http://www.pbl.nl/en/publications/opening-up-scientific-assessments-for-policy-the-importance-of-transparency-in-expert-judgements>
- The Dutch House of Representatives (2010) Kamerstuk 31793 nr. 30
- Yang Z, Jun M, Peterson RT (2004) Measuring customer perceived online service quality: scale development and managerial implications. *Int J Prod Manag* 24(11):1149–1174. doi:10.1108/01443570410563278