

REPORT OF THE COMMITTEE OF VISITORS, 2005-2007
Surface Earth Processes Section (EAR/SEP):
Hydrologic Sciences (EAR/HS)
Sedimentary Geology and Paleobiology (EAR/SGP)
Geobiology and Low-Temperature Geochemistry (EAR/GG)
Geomorphology and Land Use Dynamics (EAR/GLD)

INTRODUCTION

The Committee of Visitors (COV) for the Geobiology and Low Temperature Geochemistry (GG), Geomorphology and Land-Use Dynamics (GLD), Hydrological Sciences (HS) and Sedimentary Geology and Paleobiology (SGP) programs reviewed proposal e-jackets, analyzed program data, and spoke with Program Officers (POs) and administrators on June 2–4 2008 at NSF headquarters.

The 2008 COV members were: Charles Marshall, Chair (as a member of AC/GEO) (Harvard University), William Dietrich (University of California, Berkeley), Francisca Oboh-Ikuenobe (Missouri University of Science and Technology), Jeffrey McDonnell (Oregon State University), Kaye Savage (Vanderbilt University), David Long (Michigan State University), and Isabel Montanez (University of California, Davis).

The two charges to the COV were: 1) Review actions taken by the Geobiology and Low Temperature Geochemistry (GG), Geomorphology and Land-Use Dynamics (GLD), Hydrological Sciences (HS) and Sedimentary Geology and Paleobiology (SGP) programs during the last three fiscal years (2005-2007). Specifically, review the integrity and efficiency of processes used to solicit, review, recommend, and document proposal evaluation and actions, including the effectiveness of the program's use of NSF's two merit review criteria and to assess the relationship between decisions and program goals. 2) Evaluate the products and contributions supported and overseen by the programs over the last three years.

EXECUTIVE SUMMARY OF THE COV

Preamble: The Surface Earth Processes (SEP) Section went through a major reorganization in 2004, with the splitting of the unwieldy Geology and Paleontology Program into three programs: Sedimentary Geology and Paleobiology (SGP); Geobiology and Low Temperature Geochemistry (GG); and, Geomorphology and Land Use Dynamics (GLD). This is the first full review since this splitting of the GP program. While much of our analysis centers on the running of each of the four programs under our purview, some of our more substantive findings center on the effectiveness and impact of this re-organization.

Primary Finding (1): Outstanding Program Officers (POs). The COV was impressed by the energy and professionalism exhibited by the POs. Three distinct tasks are undertaken:

- 1) Each PO handles an enormous proposal load, with roughly half the staffing per proposal as OCE and ATM.
- 2) The POs also spend a great deal of effort in seeking co-funding opportunities, both from other programs at NSF, as well as from other agencies.

3) The POs also put in extraordinary efforts not just communicating with their communities and with other related communities, but also working to help build more cohesive community identity and structure. We wish to highlight this effort, not only because of its value in enhancing the effect of NSF on the scientific community, but also because these efforts are not readily visible in the standard metrics by which POs may be evaluated (such as the number of proposal they handle, or the dollars at their disposal, etc.).

We found that the POs handle these tasks extremely well, and the NSF and the communities they serve are very fortunate to have such a dedicated and capable team that work well together and with imaginative determination.

Primary Finding (2): Transformative effect of the splitting of Geology and Paleontology into its three daughter programs (GG, GLD, SGP). We were struck by the vitality of the three new programs, and by the new science directions enabled by their creation. When grouped together in the past, the different (even if overlapping) cultures we suspect inhibited progress. Now that each is in charge of its own future, we saw an increased vitality, with transformative effects, both in new developments under the umbrella of each program, as well as in the kinds of science each contributes to the larger whole. Here we have an interesting case of narrowing (perhaps more properly a focusing) leading to deepening and broadening. The splitting of the Geology and Paleontology program has been deeply generative.

Primary Finding (3): The programs evaluated have produced outstanding results. Over the broad spectrum of the science endeavors, including frontier science, the funded research has had a high volume, high quality, with large numbers of high impact publications (e.g., *Science/Nature* papers).

Primary Finding (4): By any metric there is a severe lack of funds. This is reflected in the three interrelated metrics:

- i) Funding Rates: All programs have very low funding rates. Over the three years of our charge the average funding rates were 13% (GG), 17% (SGP), 17% (HS) and 13% (GLD).
- ii) Average Duration of Awards: The average duration of the awards has (generally) decreased over the three years, with average for each of the four programs ranging from 2.5-2.9 years in 2005 to 2.1–2.7 years in 2007.
- iii) Average Size of Awards: The average award size in real dollars has steadily decreased, as award sizes have remained flat or decreased slightly over the three years.

Primary Finding (5): That grant sizes and durations are reaching the threshold of viability for doing effective science. There was strong sentiment among the COV members that the grant sizes are reaching the limit of the effectiveness of doing meaningful science. While this assertion is hard to quantify, we note that the average award size is approaching the cost of funding a single student, a reasonable measure of an effective minimum grant size. We are at that level now.

This descent to the threshold of viability is made even more acute by the fact that many of the PIs that go to the SEP program do not have anywhere else in NSF to seek funds, nor can they be funded by other agencies. That is to say, NSF is THE lead funding agency for many of the critical activities that fall under SEP's charge, and it should be a

high National priority to protect and preserve the science that is flourishing within the SEP section.

Primary Finding (6): Importance of SEP-funded science to critical societal issues, and the role of NSF in scientific outreach. The COV was struck by the relevance of much of the research conducted in the SEP section to societal issues, especially those related to the impact of humans on the Earth system. We were also struck by the contrasting fact that the Geosciences is considered such a low priority in OMB's proposed funding increases under the American Competitiveness Initiative. We note the important role that NSF could play in resolving this disconnect, where its nascent efforts in outreach could potentially have a major impact in communicating the societal importance of the geosciences, including the prevention or adaptation to human-induced changes, including those relating to climate, land use, and the environmental health of the biosphere and hydrologic resources.

ACTION ITEMS

Program-specific recommendations and concerns:

1) Enrichment of Hydrological Sciences: HS needs to do more to encourage PIs to generate proposals that address key theoretical areas within HS, to go after fundamental challenges in the field, and to forge new frontiers of hydrological science.

2) Need for effective mechanism(s) for achieving enrichment of HS: SEP needs to develop a strategy for achieving the recommended enrichment outlined above; this could entail changing the wording of the solicitation for proposals, to the commissioning of an NRC panel to examine the possibility of splitting the HS program (see section A.3.10). Part of the motivation for raising the possibility of splitting HS is the transformative effect of the splitting of the old Geology and Paleontology program.

3) Response to potential retirement of the HS PO: The COV is concerned with what will happen when the current long-serving and extremely capable HS PO retires. We recommend exploration to begin now on the PO post-retirement replacement, so as to minimize the potential loss of institutional knowledge and connectedness to other programs inside and outside of EAR with this retirement.

4) SGP and NCAR: Now that a 'deep-time' paleoclimate liaison has been established at NCAR, mechanisms need to be explored for making this position fully accessible to the community (e.g., funds for visiting scientists, post-doctoral fellows, students to be in-residence, workshops, on-line mini-courses, etc.). This may well take more resources.

5) SGP and the EarthTime initiative: The COV is concerned that the momentum created by the EarthTime initiative may fizzle out if further plans are not put in place for both sustaining the effort and broadening its impact.

6) HS and WATERS network: We would like clarification on the relationship between HS and the newly formed WATERS network. It is unclear if/how HS will fund the WATERS network and how this may impact HS core funding.

7) Evaluation of CUAHSI: We were not asked to evaluate CUAHSI, but given its importance and apparent success, we feel it should be evaluated. However, the COV was not provided with sufficient materials to make the evaluation.

8) GG (and all of SEP) and Broader Impacts: While the justification for funding is very well documented, there was a perception in the COV that the broader impacts criterion is not always applied uniformly, and there was concern that an appropriate balance is not always achieved in the decision-making. Discussion of this observation led the COV to raise the issue of the checks and balances that might be placed on how the criterion is applied across all of SEP.

9) GG (and all of SEP) and transformative research: While we applaud the support of high-risk projects, especially in GG, we are also concerned that some high-risk projects were funded on the basis of confidence in PI capabilities, despite potentially fatal flaws revealed in mail reviews and the panel summaries (we noted one such case in GG). While the PO provided thorough rationales for these decisions, they have the potential to undermine the competitive proposal process unless a tangible metric is defined to ensure consistency. As in action item #8, discussion of this observation led the COV to raise the issue of the checks and balances that might be placed on how to best support this type of research in SEP.

10) SGP and orphaned ESH proposals: We note that excessive dwell time for 2006-07 may reflect the fate of proposals caught in the redesign of the ESH program into P2C2. Effort needs to be made to assure that this was a one-time anomaly in processing.

11) GLD needs a permanent PO: GLD is the only program in EAR without a permanent PO, which condemns the program to constantly revisit the steepest part of the learning curve with each rotator. This impedes its maturation and weakens its ability to compete for funds (see #12 below). NSF needs to commit to a permanent PO to the program.

12) GLD's budget: We note that GLD has the smallest budget, and that in its first three years it grew less than (the poorly funded) GG and SGP. This program is attracting proposals of the highest caliber from a growing cadre of young scientists. Furthermore, GLD is at the heart of the Critical Zone, now being recognized for both its richness scientifically and its essential relevance to society. GLD needs support. We wish to commend EAR for infusing funds into the program this year, and hope that this fledgling program will be given the funds it needs to fly.

SEP-wide concerns

13) Staff (and space) needs: It is clear that SEP is still under-administered (whether measured by the proposal load/program officer, or by the excessive dwell time for the reporting of (mostly) declines). More staff is required.

14) Continued recruiting of minority-involvement proposals, reviewers and panelists is warranted. While women award rates are acceptable, the absolute numbers of women PIs is not. Neither the award rates, nor absolute numbers of minority PIs is acceptable.

15) Need for better tracking of minority groups: It came to our attention that some POs are not always aware of which submissions represent minority proposals. More attentive behavior might increase the funding rates for minorities in some programs.

16) Tribal College involvement: While steps have been taken within SEP to increase tribal college involvement, these efforts have not yielded tangible results. We hope that these initial efforts remain in place and will lead to tangible results.

17) Importance of PI's prior support: We note that, particularly in GG and HS, the POs take into consideration the PI's prior support and/or reviews of prior versions of the proposal. It would be helpful to know how this is done exactly.

18) Quality of panel summaries: While most of the panel summaries are thorough, some lack detail, in particular with regard to the broader impact criteria. More uniform attention is needed, especially given the lingering confusion over this merit criterion.

19) Satisfying the broader impacts criterion: It appears that PIs may be penalized by reviewers who pay more attention to broader impacts; therefore reviewers and panels should be made aware of the relevant statistics (see A.1.2(a) below). POs should raise the issue at town hall meetings and in their review requests, with the specific goals of communicating the importance of the broader impacts criterion, what it includes, how much effort is expected, and how to fairly review it.

20) Success rates of re-submissions: Given the low award rates it would be helpful to understand the success rate of re-submitted proposals that respond to the prior reviews, and how the panel and PO reviews help in strengthening the proposals. These statistics also need to be explicitly communicated to the community, to help PIs decide whether or not to resubmit proposals, which may help decrease overall proposal load.

Concerns we suspect are being attended to, but that we felt compelled to mention:

21) Permanent and rotator POs: The programs need to maintain the balance of permanent and rotating POs, so that they have the benefit of the historical continuity and stability that permanent POs bring, with the fresh views and energy of rotators. Importantly as mentioned in (11) above, GLD must eventually have a permanent program officer.

22) Critical Zone Observatory (CZO) sites: CZO is an important cross SEP program that deserves more attention. The COV would like more explicit information on the selection process and progress on establishing the selected observatories, as well as future funding plans for new observatories.

23) Need to moderate enthusiasm for new efforts with realism: While we applaud PO efforts' to establish new large-scale community programs, the failure of the Hydrologic Observatory and Berkeley Synthesis Center highlights the importance of tempering growth activities with realistic assessments of the challenges involved. Precautionary action in anticipation of difficulties is recommended (we appreciate that this may be difficult).

24) Importance of reviews that do not provide substantive comments: We recommend (if it is not already done) that POs either discount or down-weight reviews that do not provided substantive comments, and that these reviewers be asked to provide such comments. It would be helpful to have POs notate the e-jackets when reviews of this kind are received.

Recommendations for next COV:

25) Larger-scale initiatives: While most of the research activity within SEP was centered around standard proposals, the COV noted that some of the most important and potentially transformative efforts center around large-scale projects, such as CUAHSI and initiatives such as EarthTime. These are also often very expensive, so they need special scrutiny for that reason to. To be most effective the COV needs to hear more explicit evaluations of these initiatives from the POs during their visit to NSF.

26) Are PIs being informed of proposal decisions in a timely fashion? While the formal dwell time data indicates a severe crisis in PI notification, it appears that PIs are being informed of the decisions informally in a timely fashion. To be assured we need quantitative data on when PIs receive informal notification, data that is not presently readily available.

27) Effects of excessive dwell times: We need to know whether the slow data entry into the NSF system has any negative impact on both funded and unfunded PIs.

28) Reporting bias introduced by excessive dwell times: Except for an insert with Table B, the data sheets almost always over-estimate the award rates because many of the declinations were not entered into the system (especially for 2000). If the dwell time problem is not solved by the next COV this fact should be noted explicitly on the provided data sheets.

29) Medians versus means: We were provided with averages for award sizes, award durations, and proposal scores. However, the underlying distributions are left skewed, so medians would be more meaningful than the means. It would be helpful to either be given the median, or the underlying frequency histograms, for these metrics.

30) Funding rates between first submittals and resubmissions: To better gauge the sources of the dramatic increase in proposal loads, it would be useful to have information on funding rates between first submittals and resubmissions (see also action item #20).

31) Success of new investigators: We would like to know: 1) what proportion of successful new PI proposals were second submissions (has feedback from initially declined submissions led to successful follow-up submissions?); and, 2) how many times new PIs were funded despite a lower absolute ranking in the panel (to what extent are inexperienced PIs assisted in this way?).

32) Documenting outreach and education activities beyond the university-level: These activities appear to be increasing in response to the broader impact criterion. If NSF desires feedback on the degree to which these ideas are implemented then annual and final reports need to be made available to the COV.

33) COV consultation with POs: Finally, we recommend that future COV chairs (if not all COV members) take advantage of the time they have while “floating” to talk informally with each of the Program Officers, to get a feel for the issues that might be brewing that might not be aired in their formal reports, or in the data the COV will be provided with.

NSF-wide concerns:

34) Errors in the reporting of the number of ad hoc reviewers and proposal scores: We discovered that often either entire panels, or those panelists assigned to a proposal, are being counted as ad hoc reviewers. Similarly, we found that if an ad hoc reviewer did not offer a numeric score, or elected to report a split number (e.g. very good/good), neither the review, nor the score was recorded. These recording errors need to be fixed.

35) Alarm at pressure to reduce the number of incoming proposals: The COV was alarmed when it learned that one of NSF's management's suggestions to deal with the increased proposal pressure is for POs to try and reduce the number of proposals submitted. It is not NSF's job to try and curb the Nations scientific creative engine!

36) Education and Outreach: We feel strongly that NSF as an organization is well positioned to have major impact on education and outreach. NSF as an agency should find effective ways of publicizing the amazing range of exciting science that it funds, as well as the spectrum of innovative community outreach activities that result from NSF funded research projects.

37) Broader Impacts vs. Intellectual Merit Criteria: Like the previous COV, we are still concerned that the meaning, evaluation, and use of the Broader Impacts as a criterion for funding decisions is not well understood, or uniformly applied within and between programs. As the GG program moves to explicitly give equal weight to the intellectual merit and broader impacts criteria, the COV was moved to question whether the intellectual merit criterion is under-weighted in comparison to the broader impacts.

We note that a particularly effective way that PIs can have a broader impact is to partner with science museums, and other organizations and programs already set-up to communicate science widely, rather than have individual PIs re-invent wheel themselves.

38) Meaning of the Transformative and Interdisciplinary Research (FacTIR) criterion: While the FacTIR criterion was not yet in place in 2005-2007, we are concerned about how this will be understood and applied. We strongly urge to the NSF to continue to develop a strategy for making sure the definition is sensible, easily understood, and easily applied.

We now turn to the detailed evaluation of the programs.

PART A. INTEGRITY AND EFFICIENCY OF THE PROGRAM'S PROCESSES AND MANAGEMENT

In the assessment that follows, the Committee relied on: 1) the presentations provided by the POs; 2) information made available to us in our data folders and upon request by the POs and other staff within the SEP division while we were visiting NSF; 3) the detailed examination of a subset of the e-jackets made available to us, from 2006. Some 78 jackets were examined, drawn approximately equally from the four programs. These were divided approximately equally between awards and declines. We had hoped to concentrate on proposals that fell in the vicinity of the award/decline divide, but there was no easy way to determine which proposals fell in this category. Those examined that fell close to that boundary were given extra scrutiny.

Finally, the committee possessed considerable scientific expertise and many members have had extensive experience with one or more of the program under review, including serving on panels. This professional expertise proved invaluable in the interpretation of the data provided to us.

[CAVEAT: for the questions that involve comparison of success rates for subsets of PIs, (A.3.7, A.3.8, A.3.11, the absolute values we provide are almost all over-estimates, often to a considerable degree, because the data provided to the COV was based on the statistics in the NSF system, which does not include a sizeable fraction of the declinations (due to excessive dwell times)]

A.1 Questions about the quality and effectiveness of the programs' use of merit review process.

QUALITY AND EFFECTIVENESS OF MERIT REVIEW PROCESS

1. Are the review methods (for example, panel, ad hoc, site visits) appropriate?

As the review of proposals is at the heart of NSF's mission, and each program has slightly different needs and communities it serves, we here treat each of the four programs separately. However, we note three generalities. 1) We found the review methods for each program both sound and appropriate, and were impressed by the fact that the POs were engaged in a rigorous and explicit analysis of the proposals and reviews received, going well beyond simply tabulating average scores. We feel it is the POs job to exercise this sort of judgment, and generally feel that this was well documented, and well justified. 2) We were impressed by the POs' efforts to secure additional funding outside their programs. 3) All the POs share the difficulty in finding sufficient numbers of ad hoc reviews (in 2007 the reviewer response rate was between 46–52%), and sometimes they have to go to heroic efforts to obtain sufficient reviews (in the most extreme example, 54 reviewers (including 9 panelists) were approached to review an HS proposal).

HS: The review methods are sound and appropriate. The COV was impressed with the care and nurturing of proposals within HS. The review methods and procedures for completing reviews, panel assessments, etc., appear very rigorous and appropriate. Reviewer numbers varied greatly for the 20 proposals examined—as many as 7 to as few as 3. Even wider was the variation in reviews solicited—sometimes as many as 18 or more. In several cases, the key people in the field (for that proposal topic area) were asked to review a proposal but declined or failed to provide a review. In one of the reviewed e-jackets, the 3 reviews received were from individuals without much background in the field or standing in the discipline. The HS PO moderates the final ranking of proposals by the mail reviewers and the panel, exercising his own judgment when assigning a final ranking to each proposal. The HS PO has been adept at finding additional funding for HS-related proposals. Dollars are stretched in creative and useful ways.

GLD: GLD was a new program with a rotator who, though energetic, undoubtedly was learning how to obtain the most science with the limited funds in the program. The PO handled about 125 proposals per year. Review methods relied on both mail reviews (~4 to 5 per proposal) and panel analysis. Panelists were usually provided with outside reviews and rankings before panel meetings. Panelists read and ranked all the

proposals independently, and comparisons were made with ad hoc reviews. Discussions regarding new PIs, diversity, institutions and other relevant factors then contributed to relative ranking. Panelists made recommendations regarding relative ranking only, and the PO made the final decisions. Often the PO would seek to extend the support by seeking co-sponsorship from other panels, and by reducing budgets as deemed possible. The review methods were fair and thorough.

SGP: Overall, the review process over the 3-year period was rigorous and judicious involving both ad hoc reviewers and panel assessment. The SGP program officers were quite diligent at soliciting high-quality reviews from scientific experts relevant to the proposed research. Despite a significant increase in ad hoc review solicitations (and associated workload for the POs), the average return rate (e.g., 52% in 2007) remains good resulting in 4 to 5 reviews per proposal. No trend was observed between success rate of proposals and number of ad hoc reviews or whether the 'broader impact' criterion was evaluated by the ad hoc reviewers. The panel summaries are generally insightful and comprehensive. The PO analyses were consistently very thorough and document additional levels of independent evaluation built upon - but not solely reflective of - mail reviews and panel discussion.

GG: Overall, review methods are appropriate. The PO enlists co-review with other programs and divisions when proposals are interdisciplinary. The panel summaries and PO analyses show that this program, in an effort to support transformative research, gives preference to mail or panel reviews that point out the potential benefits of risky research, rather than always striking a balance between mail and panel reviews. This sometimes leads to funding proposals based on prior success of the PI more than the quality of the specific proposal under consideration. While the PO provides a thorough rationale for these decisions, it has the potential to undermine the competitive proposal process unless a tangible metric is defined to ensure consistency.

2. Are both merit review criteria addressed

Similar to the previous COV, we are concerned with the evaluation of the Broader Impacts criterion, and the role the criterion plays in funding decisions. Is it effectively too important? The GG program appeared to be the most aggressive in enforcing the Broader Impacts criterion, and this year this program has been explicitly weighting Broader Impacts according to NSF official decree of 50%. While we did not devote too much time to this issue, there was a strong sense that this seems to be too heavy a weight. This is an NSF-wide issue, rather than a GG, SEP, or EAR issue.

a) In individual reviews?

Based on the examination of the e-jackets, we found that the number of reviewers that explicitly addressed both criteria was variable, ranging anywhere from 50–100%, depending somewhat by program (for the proposals examined, SGP was at about 90%, HS was closer to 50%).

Alarming, we found a potential negative bias towards proposals when reviews address the broader impacts. For example, in GG, we found over the 3 years that on average awards had 3.5 reviewers address both criteria, while declines had an average of 4.4 reviewers address both. For HS, awards had 4.5 reviewers address both criteria, while declines had an average of 5.2 reviewers address both.

There are two possible explanations for this observation. First, poorly conceived broader impact plans might be gaining greater notice and commentary than typical broader impacts, in which case the increased decline rate when broader impacts are explicitly reviewed is expected and justified. Alternatively, it is possible that proposals with more ambitious broader impact goals, that also lack documentation deemed sufficient by the reviewers, receive more negative evaluations. We suspect this also happens, and feel strongly that PIs should not be penalized for ambitious broader impacts unless they are truly unachievable with available resources. This issue illustrates the need for a consistent evaluation approach for broader impacts.

Given the possibility that PIs may be penalized by reviewers who pay more attention to broader impacts, reviewers and panels should be made aware of this statistic and the POs should seek to raise this issue at town hall meetings and in review requests to communicate the importance of both what the broader impacts criterion includes, how much effort is expected, and how to fairly review it.

b) In panel summaries?

For the 78 e-jackets examined, both criteria were always addressed in GG, GLD and almost always in SGP. In HS, sometimes panel did not explicitly address the broader impacts but communicated their appreciation of the broader impacts by their enthusiasm for the project.

c) In Program Officer review analyses?

There was strong agreement that for all SEP programs the POs provide outstanding review analyses. The POs provide detailed comments of high quality, that extract key information provided by the reviewers and panel assessments; the comments are not just “cut and pasted” from the panel summary but are rewritten to synthesize the material and convey it to the PI in a nurturing and informative manner.

In the GG program there is now (but not during the period 2005–2007) an explicit effort to enforce the broader impacts criterion, giving it more weight than has been traditionally the case. There was a sense among the COV, however, in perusal of the e-jackets, that the broader impacts criterion was not always uniformly applied.

3. Do the individual reviewers provide substantive comments to explain their assessment of the proposals?

On the whole, the committee was impressed by the thoroughness of the ad hoc reviews. Many reviewers not only provide substantive assessments of the intellectual merit of the proposed research, but also provided feedback on the specific hypotheses being tested, research objectives and strategies. Often the reviews also provided concrete suggestions for improvement. Generally speaking, thorough reviews were obtained for awards that were funded as well as those that were ultimately declined. There was only minor heterogeneity in the quality and nature of the reviews across the four programs. Our sense was that the reviewers for GLD were particularly supportive (despite the low funding rate). The frequency of reviews that did not provide an explanation for the rating given was a little higher (perhaps 10%) in HS compared with other programs (although review of a larger sample of e-jackets might not support this conclusion). We recommend (if it is not already done) that the POs either discount or down-weight reviews that do not provided substantive comments, or that the reviewers be further

contacted to request such comments when they are not included with the initial review submission.

4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)?

The panel summaries reflect excellent in-depth and objective analysis by panel members and offer feedback beyond that provided by the ad hoc reviews. Generally, the panel summaries provide a clear rationale for any differences between their own assessment and the ad hoc reviews. A small percentage of the panel summaries could be more thorough, in particular with regard to the broader impact criterion, especially given the lingering confusion as to the meaning of this merit criterion.

5. Does the documentation in the jacket provide the rationale for the award/decline decision?

The committee was very impressed by the quality of the documentation. The POs have been very thorough in providing clear rationales for award decisions, particularly when there was disagreement between ad hoc and panel reviews.

6. Does the documentation to PI provide the rationale for the award/decline decision?

The information conveyed to the PIs is outstanding. The POs go above and beyond the call of duty in communicating results of the review and panel summaries in a supportive way. In several cases POs encouraged failed proposals to resubmit. The PO provides a summary of strengths and weaknesses and suggests means of improvement above and beyond the panel summary. The comments are not just taken a-critically from the panel summaries but are synthesized from the reviews and panel summaries.

7. Is the time to decision appropriate?

There is a distinction to be made between when the PIs are informed of how their proposal has fared, and when this information is formally recorded in the NSF system, the dwell time. Unfortunately, the only quantitative information provided to the COV are the dwell times. Ideally both should be less than 6 months. The reported dwell times vary greatly with the program, and in some cases, year by year:

For HS only 55-65% of proposals had a dwell time of less than 6 months, 28-41% had a dwell time of 6-9 months, and >5% had dwell times greater than 9 months. The dwell time has grown steadily worse over the 3-year review period. In contrast to the other programs, the average dwell time for awards was longer than for declines (7.1 months compared with 6.3 months). For HS the increase in dwell time may be attributed to the increase in the number of proposals.

For GG the dwell time has improved somewhat over the 3 year review period, but is exceptionally poor, with only 13% in 2005, and a still small 40% in 2007, having the target dwell time of less than 6 months. In 2006 more than 50% of proposals had a dwell time of greater than 12 months. In 2007 this number had shrunk to a still large 15%. The average dwell time for awards was 8.1 months, for declines 10.8 months.

For SGP the dwell time has also increased steadily, with 63% of proposals having a dwell time of less than 6 months in 2005, dropping to just 23% in 2007. In 2007 54% of proposals had a dwell time of 6-9 months, with 5% having a dwell time of more than a year. A subset of proposals was very delayed (21 months). The average dwell time for awards was 6.8 months, for declines 7.4 months. We note that for the period 2006-2007 the excessive dwell time may reflect the fate of EAR 'orphan' proposals caught in the redesign of the ESH program to P2C2. Effort needs to be made to assure that this was a one-time anomaly in processing (an anomaly that certainly had a detrimental effect on the efficiency and morale of the scientific community involved).

GLD has the most erratic pattern, with only 5% of proposals with a dwell time of less than 6 months in 2005 (due to the start up of the program and a new PO), with a big improvement to 75% in 2006, and then just 13% in 2007. In 2007 29% had a dwell time of over a year, when the average dwell time for accepts was 7.9 months and 10.5 months for declines. These values are too long, but apparently reflect two things: 1) the excessive workload on the sole PO who was also very active in community development projects, and 2) a tendency to hold on to projects without rejection with the hope that other funding could be found. The COV noted that the award letters were greatly delayed, in some cases over a year, while rejections were more quickly delivered.

While the formal dwell time data might indicate a crisis, it appears that PIs are generally being informed of the decisions informally well before the decisions are entered into the NSF system. It also appears that in some cases that informal decision notification was delayed while the POs sought alternative funds from elsewhere, and it appears that the relevant PIs were made aware of this when this occurred. Thus, in terms of PIs being informed in a timely fashion of the award decision, it is quite possible the reality is actually much brighter than the dismal picture painted by the dwell time data recorded in the NSF system. To be assured we need quantitative data on when PIs receive informal notification. We also need to know whether or not the slow entry of data into the NSF system has any negative impact on the PIs, both the funded and unfunded.

Finally, we note that the dwell time data make it abundantly clear that the SEP POs have insufficient staff to perform all their duties, and that NSF needs to rectify this situation. It is also a testament to the quality of the POs that their response to insufficient staff has been to delay action on secondary responsibilities such as recording funding decisions, rather than cut into the primary work of gathering first rate reviews, building new community efforts, seeking funds from outside their program, etc. While the staffing for SEP has increased in response to the last COV, their proposal load is still well above the Geosciences Directorate average, and the section is still under-staffed.

8. Additional comments on the quality and effectiveness of the programs' use of merit review process:

In sum, we found the POs highly engaged and proactive in review solicitation, in assessing the results of the mail reviews and panel, and in communicating with the PIs. Given the low award rates, and all the effort the POs are putting into the review process, it would be helpful to understand the success rate of re-submitted proposals that respond to the prior reviews, and how the panel and PO reviews help in strengthening the proposals.

With regard to the growing emphasis on funding transformative research, we note that, particularly in GG and HS, the POs take into consideration the PI's prior support

(compliance with reporting) and/or reviews of prior versions of the proposal. It would be helpful to know how this is done exactly.

As a final comment, it is obvious that running an NSF program is a difficult and complex task, requiring good judgment and advice, and that the POs are doing an excellent job in applying this judgment to best achieve their goals.

A.2 Questions concerning the selection of reviewers.

SELECTION OF REVIEWERS

1. Did the program make use of reviewers having appropriate expertise and/or qualifications?

The POs displayed considerable skill in selecting reviewers and often selected leaders in the field. Generally speaking, reviewers were consistently appropriate and of high caliber. However, reviewer declines sometimes frustrated the POs concerted efforts. More frequently in the GG and HS programs the key people in proposal topic area were asked to review a proposal but declined or failed to provide a review. In one case of the 20 e-jackets examined, the 3 reviews that were received were from individuals without much background in the field or standing in the discipline. In response to the difficulties in finding appropriate reviewers, GG panelists were sometimes solicited to provide reviews.

The POs also reached outside geosciences at times to ensure that proposals at the edge of the discipline were properly reviewed.

2. Did the programs use reviewers balanced with respect to characteristics such as geography, type of institution, and underrepresented groups?

For each program, the geographical distribution of reviewers was well balanced and roughly corresponded to the distribution of proposal origins.

Reviewers were heavily weighted towards Ph.D.-granting institutions, especially research-intensive Ph.D. institutions. This latter group is also disproportionately represented by a larger number of submissions. The variance between programs was small, with approximately 19–21% coming from Ph.D.-granting institutions, 32–36% from research-intensive Ph.D. institutions, with about 23–31% of reviewers not reporting the type of institution to which they belong. Only 2.5% or less of reviewers were from 2 or 4 year institutions.

Developing statistics for the balance of reviewers from underrepresented groups is made difficult by the fact that anywhere between 75–81% of reviewers did not report their gender, whether they belong to a minority group, or have disabilities. Of those that reported gender, anywhere from ¼ to ½ were female, one in ten (or less) were a minority, and one in 13 to 18 reported being disabled. Perusal of the e-jackets suggests a distribution of gender and ethnicity, and disability in the ad hoc reviewers is reflective of the community.

3. Did the programs recognize and resolve conflicts of interest when appropriate?

As far as we can tell, the programs recognized and resolved conflicts of interest when appropriate. Based on examination of the e-jackets, typically 3-5% of ad hoc reviewers identified conflicts of interest (data only from LDG and SGP), suggesting the POs are doing a good job of avoiding this problem. The previous COV report mentioned that 30% of reviewers did not complete the COI. We were not given data on the frequency under this review, but we see no evidence or reason to suspect serious breaches of conflict of interest (we suspect that vast majority of those not completing the COI fail to do so because there is no COI, but we have no data to back this up).

4. Additional comments on reviewer selection:

None.

A.3 Questions concerning the resulting portfolio of awards under review.

RESULTING PORTFOLIO OF AWARDS

1. Overall quality of the research and/or education projects supported by the programs.

The COV was very impressed with the overall quality within the four programs.

The GG program has responded particularly well to new advances and emerging topics including nanogeoscience, geosciences in relation to human health, and interactions among physical, chemical, and biological processes. GG awards also cross a broad range of spatial scales. GG in particular has been funding high risk projects, but while we applaud the support of high-risk projects, we are concerned that some high-risk projects were funded on the basis of confidence in PI capabilities, despite potentially fatal flaws revealed in mail reviews and detailed in the panel summaries.

The COV found the GLD proposals to be distinctly exciting, innovative, and potentially transformative. A high standard was set at the beginning with an emphasis on hypothesis testing, process mechanics, modeling and new field observations. Education outreach was not well developed in some of the funded projects.

For the HS program the overall quality is very high but without final reports, it is difficult to assess the end metrics (e.g. published research)(this of course applies to all programs). There was a relative lack of submitted proposals with an emphasis on hypothesis testing and theory development.

For SGP the overall quality was very good. Despite the splitting of the old Geology and Paleontology program into three, the intellectual scope with the new SGP is still very broad, and a diverse range of science is being supported within the program. Informally, the program director for SGP raised the issue of whether the program would benefit from being divided into two, but the COV did not discuss this suggestion.

2. Do the program portfolios promote the integration of research and education?

Across the four programs we found strong evidence of an effort to integrate research and education. These efforts included funding of proposals with creative and well-developed educational components (including K-12 and public outreach) incorporated into the broader impacts, the funding of workshops, in addition to more traditional approaches focused on training graduate students. A small number of CAREER awards in which integration of research and education is a critical component, were also made, as well as the support of several REU proposals.

Our impression is that proposed activities involving outreach and educational activities beyond the university-level are increasing in response to the broader impact merit criteria. However, it is not possible to evaluate to what degree these ideas are implemented or carried through given the information in the e-jackets. If NSF desires feedback in this area then annual and final reports need to be added to the e-jackets and made available to the COV.

3. Are awards appropriate in size and duration for the scope of the projects?

There is considerable variance among programs and from year to year in both award sizes and durations. The three-year averages are: for GG, \$76,000/yr for 2.3 years; for HS, \$101,000/yr for 2.6 years; for SGP, \$80,000/yr for 2.2 years; for GLD, \$83,000/yr and 2.1 years, although in 2005 the average award size was \$177,000/yr for 2.8 years over 10 awards, while in 2006, presumably in the attempt to stretch the dollars, the average award size was just \$46,000/yr for 1.7 years for 24 awards.

Despite the variance sobering generalities emerge. First, award sizes and durations have either remained static, or shrunk over the three-year period, and the award sizes are not significantly different from those 10 years ago. Thus, award sizes have not kept pace with increasing costs for students, salaries, and overhead, and the shorter durations lead to more frequent proposal submission, burdening the researchers, reviewers, and NSF staff. While shorter duration is one way to stretch inadequate funds, it is not good for the system or the researchers, in particular in light of the greater expectations of integrating outreach and non-traditional activities. There is concern that a threshold of funding is being reached – one in which the true cost of research is greater than the actual award amount resulting in PIs ‘contracting’ to carry out more research than can be accomplished well at the actual funding level. Furthermore, the awards sizes (and durations) may hamper transformative research or directions that require investment into new resources. Finally we note that the situation is actually worse than it seems; the awards size distributions are generally left-skewed, so typically some 70% of the awards are likely smaller than the reported averages, perhaps in the order of \$10,000/yr. Thus, to pick an extreme case, this means the median grant size for GLD in 2006 was absurdly low, especially given that the mean of \$46,000/yr includes the indirect costs.

4. Do the program portfolios have an appropriate balance of:

- **Innovative/potentially transformative projects?**

This was not easy to judge, in part because most of the proposed projects have not yet reached fruition. This is exacerbated by the youth of three of the four programs, and the small number of grants so far awarded. One metric of innovative and potentially transformative projects is to count the number of SGER grants awarded, but we decided

that is not really a good measure of the balance of innovative/potentially transformative projects, in part due to their small number, in part because the exact rationale for their funding is not always clear, and in part because they have very specialized criteria for being awarded, including small budgets.

Turning to each program in turn, for GG, we note that there are several funded projects that address new research areas that have high potential for transformative impacts, particularly in biogeochemistry and geomicrobiology. Examination of the e-jackets and personal experience by members of the COV with GLD indicates that the GLD proposals are tipped towards the transformative. SGP has adopted a policy of using SGER grants to fund pilot or “proof of feasibility” projects as a way of fostering innovative and potentially transformative projects. Overall, in these three programs, SGP, GG and GLD, there appears to be a healthy balance of innovative and potentially transformative science. In contrast, our sense was that the HS community is somewhat risk-averse; proposals are of high quality but generally do not propose transformative work, focusing on observation and modeling, in contrast to hypothesis-driven research.

5. Do the program portfolios have an appropriate balance of:

- **Inter- and Multi- disciplinary projects?**

The POs are very pro-active and effective in seeking cost-sharing agreements with other programs, in participating in special cross-disciplinary initiatives (including initiatives managed outside of SEP), and in soliciting partnerships outside of NSF. The programs are also strongly supportive of multi-PI and multi-institutional collaborative projects. These efforts take considerable time and skill execute successfully. Each of the four programs also has a significant number of their own projects co-funded with a wide range of programs both within the EAR directorate and with other directorates (each program was able to secure co-funding from anywhere between 3-6 other directorates, and from 18 to 32 different programs). To a lesser degree, each program also co-funds proposals based in other programs. When these co-funded proposals are added to the subset of collaborative proposals funded entirely within the core programs, it is abundantly clear that there is a healthy intellectual interchange between the programs and other disciplinary units within NSF. The POs have been energetic and creative in securing funding from other programs, especially for proposals that just missed the funding cut determined by the panel.

Ideally, the next step would be to increase the duration of such collaborative efforts – this would encourage more innovative/transformative research while reducing proposal load and presumably rejection rates.

6. Do the program portfolios have an appropriate balance considering, for example, award size, single and multiple investigator awards, or other characteristics as appropriate for the program?

The number of collaborative projects awarded ranged from 24% (GG) to 25% (HS) to 32% (SGP and GLD) of the total number of competitive awards (where each project is counted as one award). In all programs, the majority of these went to 2-part collaborative proposals (94% in GG, 77% in HS, 65% in GLD and 56% in SGP). These seem reasonable values, although the quality of the projects is the paramount criterion, not their number.

7. Do the program portfolios have an appropriate balance of:

- **Awards to new investigators?**

The answer is in general is yes. All programs have a healthy number of new PIs, especially for HS and GLD. The funding rates for these typically new PIs is consistently less than PIs with prior funding, ranging from with new SGP PIs being funded at 4/5ths the rate of the experienced PIs, dropping to about 2/3rds for GG and GLD, and to just less than half for HS.

For GG submissions by new PIs accounted for 26–35% of the total proposals, with an average of 29%. Success rates (see caveat on page 8) for these proposals were 13–17%, with an average of 14%, compared with an average of 21% for the PIs with prior success.

For SGP, new PIs accounted for between 28–29% of all proposals, of which 14–25% were funded, with an average of 19%, which is only slightly less than the 3 year average of 23% for PIs with prior success.

GLD is an unusual program in that many of the PIs are unusually young. New PIs accounted for 30–43% of proposals (average 36%), of which 20–35% were funded (average of 28%), compared with an average of 43% for prior PIs for the program.

HS has an even newer PI community, with submissions for new PIs ranging from 41–50%, with an average of 44%. Funding rates were 6–18%, with an average of 11%. These are somewhat lower success rates than the PIs with prior funding, with an average of 24%.

An important metric that was not available to the COV is how many of successful new PI proposals were second submissions. It would be useful to monitor whether feedback from initially declined submissions led to successful follow-up submissions. Further, it would be interesting to see statistics on how many times a new investigator was funded despite a lower absolute ranking in the panel; that is, to what extent are new PIs given the benefit of the doubt for inexperience in writing competitive proposals?

8. Do the program portfolios have an appropriate balance of:

- **Geographical distribution of Principal Investigators?**

While in detail the pattern is complex, there is a strong correlation between the number of submitted proposals and the numbers of institutions/PIs in the states able to take part in the competition. Given low funding rates, and the fact that review process evaluates proposals based on their intrinsic merit, there is not a perfect correlation between the geographic pattern of submittals and awards (e.g., in HS there were 13 and 14 submissions in Tennessee and Connecticut with no awards), but in general states with the higher number of submittals also had higher numbers of awards. In the case of EPSCoR states, some states did not receive funding (but the total number of submissions is small), but generally speaking awards were concentrated in the centers that emphasize the research relevant to the program they were submitted to. EPSCoR funding rates (see caveat on page 8) are less than the national average to a greater to lesser degree (for HS, 12% compared with the program average of 24%; SGP, 15% compared with 22%; GG; 17% compared with 19%; GLD 34% compared with 37%), but still generally comparable (except for HS were it was half the average rate). Thus, on balance, the COV considers the geographical distribution of funded PIs appropriate.

9. Do the program portfolios have an appropriate balance of:

- **Institutional types**

Averaged over all programs the research intensive Ph.D. institutions dominate the number of awards (63%), followed by Ph.D. institutions (22%), then at much reduced levels, the Masters institutions (6.5%), Business, State & Local, Foreign institutions (6.6%), and then the 4 yr (2.5%) and 2 yr institutions (0.1%).

For GLD the dollar award sizes are in the same proportion as the number of awards. For GG research intensive institutions are disproportionately represented, accounting for 86% of the dollars, with Master programs being especially under-represented. For HS, both research intensive institutions and the Business, State & Local, and Foreign category are overrepresented, while for SGP there is also overrepresentation for the Business, State & Local, and Foreign category. These differences are in accord with the cultural differences between the various research communities represented by the different programs.

In short, the program portfolios appear to have an appropriate balance of institutional types.

10. Do the program portfolios have an appropriate balance:

- **Across disciplines and sub disciplines of the activity?**

The review of the GG, SGP and GLD e-jackets, as well as the experience of the members of the COV, suggest that these programs have an appropriate balance across disciplines and sub-disciplines. Moreover, there is also a healthy balance of hypothesis testing, theory development, and detailed case studies. For GLD, the newest program in SEP (in the sense that GG and SGP inherited experienced POs from the old Geology and Paleontology program), with community support through the panel, this program has set a very high standard. A balance between the various sub-disciplines ranging from glaciology, landscape evolution processes, eolian processes, to landuse management theory has been purposely sought (both through funded awards and panel composition).

In contrast, the COV noted that many of the projects funded in HS were based on strong case-study approaches, with a relative absence of proposals that could lead to transformational studies and high-impact publications. Unlike the proposals in the 3 other programs, HS proposals were generally not hypothesis driven or focused on fundamental theory. Very few proposals addressed pressing hydrological theory needs, although new theory is being advanced outside the core area (e.g., through the Illinois Synthesis Center, from transformative insights gained from new measurements enabled by the CUAHSI-based fiber optic temperature initiative, and through many of the workshops and training sessions funded by HS).

The COV feels that the HS program should do more to encourage PIs to generate proposals that address key new transformative theoretical areas within HS, to go after fundamental challenges in the field, and to forge new frontiers of hydrological science. One way of doing this might be an updating of the HS mission statement that is communicated to PIs during the proposal solicitation process. Community science reports from the HS funded workshops and from the CUAHSI Vision Workshops could be incorporated into the HS mission statement to help direct more forcefully new proposals that strike at this needed part of the HS portfolio--testable hypotheses to

motivate future field and theoretical research. (We note here that the COV did not have access to any of the CUAHSI or Synthesis Center details).

To fully realize the enrichment outlined above, the COV suggests the possibility of splitting the HS program. Perhaps an NRC panel could be commissioned to examine HS, and to consider a split. Giving the reinvigoration engendered by the splitting of the old Geology and Paleontology program into the 3 new SEP programs, and given that HS is the largest of the four SEP programs, as well as the need for fostering growth in the new area of theoretical hydrology, such a sub-division (timed with the HS PO retirement) could infuse the hydrological sciences with a new sense of purpose and focus. While the COV is not necessarily suggesting such a split, an NRC commission could take on such an examination, reflect on the much-needed update to the 1991 NRC Report on Opportunities in the Hydrological Sciences, and attempt to realize, programmatically, the many recent outcomes of various CUAHSI workshops and Hydrological Synthesis Center activities.

11. Do the program portfolios have appropriate participation of underrepresented groups?

There are two issues here. First, the proportion of underrepresented groups that submit proposals. Second, of those that submit proposals, the proportion funded. Gathering data to answer these questions is made difficult by the fact that many PIs and reviewers do not report their gender, or whether they belong to a minority group. For minority groups, statistical generalizations are further hampered by the fact the absolute of number are typically very small.

With regard to submission rates, we do not have the data available to determine whether women and/or minorities are submitting proposals in proportion to their representation in the profession, and in proportion to the types of institution they work for. Our feeling, without hard data, is that females and especially minorities are under-represented in the profession, and that both are probably especially under-represented in the research-intensive institutions.

With regard to the funding rates (see caveat on page 8) for proposal with women involvement, it appears that women are funded at comparable rates, sometimes a little higher, sometimes less, than the average rates for the programs (for HS, 21% of proposals with women involvement were funded compared with the program average of 24%; for SGP, 24% compared with 22%; GG, 24% compared with 19%; and GLD, with 29% compared with 37%).

To our surprise, minority groups were sometimes funded at comparable rates to program averages (for SGP, 24% of proposals with minority involvement were funded compared with the program average 22%, and for GLD 38% were funded compared with that program average of 37%), but in the other two programs they were funded at a lower rate than the average (for HS, 17% of proposals with minority involvement were funded compared with the program average of 24%; for GG, 6% were funded compared with the average of 19% for the program). However, we note the total number of submissions was so low that the statistics are almost meaningless (e.g., for HS there were just 12 minority awards over the 3 year reporting period, for SGP 8 awards, for GLD 5 awards, and for GG just 2 awards). Informal discussion with some of the SEP staff indicated that some POs are not always aware of which submissions represent minority proposals, and so more attentive behavior might increase the reported funding rates for minorities in

some of the programs. Continued recruiting of minority-involvement proposals is warranted. We also note that the effort taken by SGP to get more tribal college involvement has not yielded tangible results; but there has been some progress in getting more minority-serving institutions involved.

12. Are the programs relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports.

The answer in short is yes. The four SEP programs reviewed foster research in areas identified by a number of NRC reports, including the 2007 report “Earth Materials and Health: Research Priorities for Earth science and Public Health”, which highlights the need for “improved understanding of the source, fate, bioavailability, and impact of potentially hazardous or beneficial earth materials”, the 1991 report on “Opportunities in the Hydrological Sciences”, and the 2001 report on “Basic Research Opportunities in Earth Science”. This last report specifically called for research in the Critical Zone, with emphasis on land use dynamics, which underpins the research mission of the GLD, in strong partnership with GG. With various forms of community input, including the Workshop on Critical Zone Exploration Network, these efforts have led to the establishment of the Critical Zone Observatories (CZO). HS has also benefited from the HS/CUAHSI-sponsored Vision Workshops, but could be more proactive in adopting some of their recommendations.

The various programs also have forged strong relationships with partner organizations, such as the USGS, NASA, DOE, DOC, the US international hydrology efforts through the US Committee for International Hydrology, the Army Research Office, and inter-agency support for the Hydrologic Information Service (HIS) (via EPA, NOAA, NWS, NCDC, and OGC).

13. Additional comments on the quality of the projects or the balance of the portfolio:

None.

A.4 Management of the program under review.

1. Management of the programs.

HS: The PO is a skilled program administrator, who has guided the HS for many years and shown imagination, innovation and commitment for furthering the science of hydrology. The solicited reviews, panels and PO reports to the PIs are professional, encouraging, and nurturing. The hydrological community is very fortunate to have had someone so dedicated at the helm for so many years. One concern is what will happen when the current PO retires, with the loss of institutional knowledge and connectedness with other programs inside and outside of EAR. We recommend exploration begin now on the PO post-retirement replacement, including engaging the hydrological community in identifying their desired candidate(s).

GLD: The new program was initiated by Enriqueta Berrara, and quickly moved into action through her experienced guidance. Her knowledge was important in getting GLD set up. Mike Ellis (a rotator) then was hired and has now completed his three-year rotation. He proved to be energetic, creative, fair, and demanding (in the best sense of

the word). He stimulated the community to action, and supported grass-roots efforts. He made a significant effort to learn about the diverse fields under the GLD purview (by attending meetings, reading, and talking to many). He successfully sought alliances within NSF and beyond, and attempted to stretch very thin resources. He supported formative workshops, new initiatives (e.g. the Critical Zone Observatory), and earned the high respect of the community. The only complaint heard about his program was the severe lack of funds which forced so many promising proposals to go unfunded.

SGP: The SGP POs are to be commended for their thoughtful, objective and effective management of the program. Notably, their efforts in community outreach to help identify and foster new areas of research (in particular, interdisciplinary research and cross-fertilization of techniques and intellectual resources) and their support of 'pilot or feasibility studies' in order to stimulate potentially transformative research is highly commendable. The degree of outreach to the community is well beyond expectation especially given the high proposal load of the SGP program.

GG: The GG PO has actively engaged with other federal agencies (NASA, USGS, DOE) in joint proposal review and funding and in securing use of facilities by PIs that are funded by GG. The program has also undertaken the most ambitious steps to enforce the broader impacts criterion, to increase the proportion of proposal with women involvement, and to support high risk and transformative research. As discussed elsewhere in the report, while the justification for funding is well documented, there is a perception in the COV that the broader impacts criterion is not always applied uniformly, and at times PIs with strong track records are sometimes funded for proposals where both the ad hoc reviewers and panel have identified serious weaknesses. While we are supportive of POs taking an active role in these aspects of the funding portfolio, there is concern that an appropriate balance is not always achieved in the decision-making.

2. Responsiveness of the program to emerging research and education opportunities.

All programs have been very responsive to emerging research and education opportunities. The COV was very impressed with the energy and diligence the POs have engaged in this component of their responsibilities. Below we outline some of the highlights for each program.

GG: This program has been involved with several initiatives that tap into new opportunities, including supporting and participating in workshops and short courses (e.g. a workshop on Paleoenvironmental Proxies (0609120), an International Workshop on Bio-Soil Interactions in Engineering (0628782), and a short-Course on Medical Mineralogy and Geochemistry (0628748)). GG also supported a Molecular and NanoEnvironmental Geochemistry working group meeting (0630535), and participated in the 9/2007 "Emerging Topics in Biogeochemical Cycles" proposal call aimed specifically at interdisciplinary research. There were 10 SGER projects funded, several of which focused on paleobiology, isotope fractionation, or aspects of the carbon cycle. One addressed the effects of Hurricane Katrina on organic matter redistribution.

GLD: This program has not only been responsive, it has been stimulating the community to drive towards emerging areas. It funded 13 workshops and symposium, and 13 SGER awards during the 2005-2007 period. The PO notably encouraged the series of workshops in 2007-2008 intended to help delineate key new research areas in GLD including: Complexity in Geomorphology; Climate over Landscapes; Life and

Landscapes (MYRES III) and Studying Earth Surface Processes with High-Resolution Topographic Data. GLD funded the NRC study: Challenges and Opportunities in Earth Surface Processes (0738045), which will provide a report on the state of the field in early 2009. The PO became the joint technical coordinator for NCED. GLD developed a partnership with OCE for the co-funding of the Community Surface Dynamics Modeling System (CSDMS) and became its primary program. CSDMS will play the crucial role in the development for the first time of a community-level shared modeling capability. CSDMS has held several community workshops to initiate its program. The PO actively supported the proposal for the formation of an AGU Focus Group (Earth and Planetary Surface Processes) and made presentations at AGU and GSA meetings about GLD to introduce this new program. In total, this shows an enormous effort to build community and respond to emerging research and educational opportunities.

SGP: The COV noted the broad spectrum of workshops and related community/educational outreach efforts (e.g. EarthTime, GeoSystems, PaleoStrat) supported by SGP. Seeding across-the-spectrum efforts/initiatives is necessary initially to allow for and assess community interest/response; after an initial period though the POs are encouraged to focus on those that prove to be most promising (i.e., engage the largest/broadest faction of the community; areas of research reflective of areas in which the program is most broadly invested; unique and integrated educational-research opportunities). One such initiative that holds promise for significant impact based on a relatively small current investment, is the recent establishment of a 'deep-time' paleoclimate liaison at NCAR. This investment has the potential for integrating climate modeling into many of the subdisciplinary areas of research currently funded by SGP (paleoclimatology, paleoceanography, paleoecology, genetic stratigraphy, and paleobiology), and for training a new generation of sedimentary geologists, paleobiologists and geochemists who are truly interdisciplinary. This will, however, require additional support to bring this opportunity to fruition for PIs, students and post-doctoral fellows from the broader community, perhaps via in-residence fellowships and workshops, additional liaisons, etc.

HS: The PO has been very responsive to community needs and emerging research and educational efforts. CUAHSI has been a major thrust area in this regard. The joint HS-CUAHSI Vision workshops, instrumentation training workshops and recent CZO awards are all tangible examples of this. The test-bed proposals and synthesis center activities are also important outcomes during the period under review by this COV. One area that is unclear is the relationship between HS and the newly formed WATERS network. It is unclear if/how HS will fund the WATERS network and how this may impact HS core funding. The SAHRA and NCED programs have been huge community successes.

One concern that emerged as the COV deliberated was an issue highlighted by the HS program, and that is the large investment in time, energy, and enthusiasm in the development of large-scale community activities that have subsequently failed, such as the Hydrologic Observatory and Berkeley Synthesis Center. Perhaps this is just the price for trying to develop new broad-scale activities. But while the NSF is to be congratulated in spear-heading such initiatives, failures of this kind have a dampening effect on the communities involved, and so some attention must be paid to the long-term feasibility, in terms of personal and finances, of these activities. We have no concrete recommendations, but felt it important to report this observation.

3. Program planning and prioritization process (internal and external) that guided the development of the portfolios.

We feel that all four programs are run well, and, given the nature of their communities, have well developed portfolios. Two of the programs, GG and SGP, while new, have experienced POs, with well-established communities and practices. Both POs have taken the opportunity, now that the programs are distinct, to work with their communities to define new research directions and opportunities. The third program, GLD, was led by a new PO Mike Ellis. Mike aggressively sought advice, obtained from panel members, workshop results and individual discussions. Given the limited amount of funds he did an admirable job of building a broad and high quality portfolio. In our judgment, all three have healthy and well-developed portfolios. For HS, we encourage the PO to consider augmenting the current portfolio to include more proposals that address key new transformative theoretical areas within HS, to go after fundamental challenges in the field, and to forge new frontiers of hydrological science. One way of doing this might be an updating of the HS mission statement that is communicated to PIs during the proposal solicitation process. Community science reports from the HS funded workshops, the CUAHSI Vision Workshops could be incorporated into the HS mission statement to help direct more forcefully new proposals that strike at this needed part of the HS portfolio--testable hypotheses to motivate future field and theoretical research.

4. Responsiveness of program to previous COV comments and recommendations.

Many of the problems identified by the previous COV centered on proposal load and low funding rates, for which little can be done at the program level, although new staff were hired. For the other issues raised, we found the responses a little anemic, especially after the first year. However, the easiest to solve problems were addressed.

The prior COV identified biogeosciences as a critical developing research area, but noted that confusion arose over the variety of short-lived programs addressing relevant research topics. The GG program has now been active for three years and seems to have effectively stabilized, although the extremely high proposal pressure shows that the need is still not being fully addressed with the current funding levels, and in fact the proposal success rate has declined over the three-year period of review from 16% to 12%. The GG program manager has developed relationships with other programs that can extend its funding capability to a degree, but there remains a serious lack of sufficient funding for core awards.

The HS has responded to the last COV by improving its communications to PIs through more detailed and synthetic reviewer/panel feedback, formulation of broader impact metrics, improving the process for co-reviewing proposals, and increasing the panel composition to 9.

The SGP program responded well to the previous COV, implementing those changes feasible under budget and time restrictions.

5. Additional comments on program management:

A final component of program management not covered above is the how the panels are run. We were not given much information on this, but were impressed by the efforts in HS to improve the HS panel structure through the inclusion of a small number of new

PIs, establishing a better gender and minority balance, and including a good range of senior investigators to cover the relevant sub-disciplines. Interestingly, HS give panelists 2 weeks to write panel summaries after panel, which leads to high quality summaries. However, the other panels are also producing high quality panel summaries, so we do not feel strongly that other panels need adopt the HS approach. In SGP, there has been some panel instability, with terms of no longer than 2 years, and with many only serving on one panel. While we do not by any means want to suggest micro-management of the panels, we note that there is a need for some continuity in panelists, so that historical memory and panel standards have a degree of stability and predictability. This is an important component of the review process, especially with the rising number of re-submissions created by the shortage of funds.

The COV was struck by what might be called cultural differences between the four programs, including different ways their panels are run (see above), the types of science they support, the degree to which they seek funding from other parts of NSF as well as outside NSF, etc. We feel that each program is running well, and that these differences reflect in part the different natures of the communities they serve, and in part differences in the personalities of the POs. We feel that these differences are a healthy sign that each program is being run with the specific goals of their programs in mind.

In terms of running programs, the COV spent some time discussing the importance of having both permanent POs (who can provide historical continuity and can establish long-term relationships within NSF and the community), and rotators (who provide fresh views and energy for each of the programs). The COV considers the lack of a permanent PO in GLD as detrimental to the continued success of this new program. The current POs in the other three programs have used their considerable experience and skill to creatively guide and build their programs, both scientifically and in terms of their budgets. The continuity of leadership in GG and SGP will be beneficial to the further growth of these new programs. The first GLD PO did a superb job alerting, encouraging, and, with meager funds, supporting this rapidly growing community. This effort needs to be sustained but it seems unlikely that GLD can be successful when all leadership knowledge is lost every few years with each new rotator. This lack of continuity will also affect the current community efforts in the Critical Zone Observatories, CSDMS, and NCED, and make less likely the emergence of other such programs.

In short, all programs would be best served with both a permanent PO and a rotator, and we hope that this balance can be met in the foreseeable future.

Finally, as noted above the proposal load is very high for the SEP POs; they need help, both in terms of office space, and also with staff. We strongly recommend more staff be hired (for example, SGP only has a ¼ time assistance, and need another PO). One of the reasons the programs are running so well, is that the POs are working well beyond the call of duty in support of their programs. As just one anecdote in support of this, one of the POs commented informally to the COV that their program is lucky to have a PO who does not have kids and has a very supportive spouse; a younger PO with more family responsibilities would not be able to work as late into the evenings and over the weekends, and thus would not be able to do the required office work AND engage the community.

PART B. RESULTS OF NSF INVESTMENTS

The NSF mission is to:

- promote the progress of science;
- advance national health, prosperity, and welfare; and
- secure the national defense.

To fulfill this mission, NSF has identified four strategic outcome goals: Discovery, Learning, Research Infrastructure, and Stewardship. The COV should look carefully at and comment on (1) noteworthy achievements based on NSF awards; (2) ways in which funded projects have collectively affected progress toward NSF's mission and strategic outcome goals; and (3) expectations for future performance based on the current set of awards.

B. Please provide comments on the activity as it relates to NSF's Strategic Outcome Goals. Provide examples of outcomes ("highlights") as appropriate. Examples should reference the NSF award number, the Principal Investigator(s) names, and their institutions.

B.1 OUTCOME GOAL for Discovery: "Foster research that will advance the frontier of knowledge, emphasizing areas of greatest opportunity and potential benefit and establishing the nation as a global leader in fundamental and transformational science and engineering."

We here split our response into two sections, the first dealing with the already established program HS, the second dealing with the three daughters of the old Geology and Paleobiology program, SGP, GG and GLD.

HS: This program has fostered research that is leading the important new developments in science and engineering. For example, the recent fiber optic measurement technique sponsored by HS has created an explosion of activity in the hydrological sciences where this new approach is opening up entire new avenues for research, from detecting groundwater inflows to streams along stream corridors, to distributed soil moisture measurement. The work on separating the effects of albedo from eco-physiological changes on surface temperature is also fundamental.

The COV notes that while such CUAHSI initiative appear impressive, we were not provided with sufficient materials to evaluate CUAHSI and were not asked to do so.

However, while working at a high level, as described above, we also feel that NSF could play a leadership role in fostering the development of research programs centered hypothesis testing and theory development within the HS community.

GG, GLD, SGP: The breaking up of the old SGP program into the three new programs has had a tremendously stimulating effect, making possible entirely new research of the kind that can be truly transformative: that is, the narrowing has led to broadening. It has allowed those working at the boundaries of various disciplines to make innovative proposals and for mail reviewers and panelists perceptive of new directions to applaud rather than reject these new efforts. As part of the revitalization the new programs have engendered, the POs for these programs have focused on community building through workshops, town hall meetings, shared infrastructure and cross-disciplinary initiatives.

It is difficult here to select a few “discoveries” to highlight here. Part of the problem is that most of the projects funded under the period of review are still underway or just moving to publication. Programs with a deeper legacy (SGP) can rely on papers that arose from earlier funding but were published during this period. In this case we note that SGP reported 41 papers in *Science* and *Nature* published between 2004 and 2006. Each program provided the COV with example discovery papers and collectively these works are impressive in their quality, significance and range. The COV did not attempt to derive a separate analysis of the publications derived from funding during this period, as this seemed to be beyond our mandate.

B.2 OUTCOME GOAL for Learning: “*Cultivate a world-class, broadly inclusive science and engineering workforce, and expand the scientific literacy of all citizens.*”

Beyond the critically important training of post-docs, graduate students, and undergraduates directly through the research supported by NSF grants, and through the enhanced teaching that flows from the involvement of College and University instructors in cutting-edge research, the four SEP programs have been involved in a wide range of effective activities that enhance the cultivation of a world-class, broadly inclusive science and engineering workforce, and, to a lesser extent, expansion of the literacy of all citizens. The four programs are all involved in the development of short courses, the support of CAREER grants, REU programs (to cite just one of many examples, the Environmental Molecular Science Institutes), and, within HS, the CUAHSI program. The POs have been exemplary in reaching out to the scientific community through town-hall meetings at National meetings, and through a broad range of interactions with PIs and potential PIs. By and large the impact on expanding the scientific literacy of all citizens has played less of a role, but SGP’s and other programs involvement in NSF podcasts points the way for NSF to be much important in conveying the excitement and importance of the discoveries made with NSF funds to the greater public. The COV strongly feels that NSF can and should play a leadership role in communicating the excitement, importance and need for geosciences in the National and International arenas.

B.3 OUTCOME GOAL for Research Infrastructure: “*Build the nation’s research capability through critical investments in advanced instrumentation, facilities, cyberinfrastructure and experimental tools.*”

Here too, the SEP programs have made significant advances in basic research infrastructure. Spectacular successes have been the CUAHSI instrumentation & informatics programs, and the establishment of NCED, co-managed by SGP and GLD, CZO and CRONUS. In addition, EAR IF also supports programs that are particularly relevant to SEP, including NCALM and PRIME lab.

While we understand that the scientific goals of the CZO sites maybe quite different from the already established and successful LTER sites, we would hope that an effort is made to leverage the existing LTER sites, if scientifically sensible, when establishing the CZO sites.

PART C. OTHER TOPICS

C.1. Please comment on any program areas in need of improvement or gaps (if any) within program areas.

Our primary concern here is the need for increased numbers of proposal centered on hypothesis testing and theory development in HS (as noted above).

C.2. Please provide comments as appropriate on the program's performance in meeting program-specific goals and objectives that are not covered by the above questions.

No further comments.

C.3. Please identify agency-wide issues that should be addressed by NSF to help improve the program's performance.

First, there is serious concern over the low number of awards, the small award sizes, and the shortening duration of awards, especially given the large numbers of high quality proposals that are not funded. We are particularly concerned that the award sizes and durations are reaching the threshold of effectiveness.

The COV was seriously alarmed when it learned that one suggestion from NSF management to deal with the increased proposal pressure is for POs to try and reduce the number of proposals submitted. It is not NSF's job to try and curb the Nation's scientific creative engine, and we were very pleased (and relieved) to see that the POs have grabbed the bull by the horns by dealing creatively, and with a remarkably high morale, with the increased proposal load and decreased funding rates.

We feel strongly that NSF as an organization is well positioned to have major impact on education and outreach, and we feel that NSF, the sciences, and the Nation as a whole, would benefit greatly if NSF as an agency could find effective ways of publicizing the amazing range of exciting and diverse science that it funds.

Like the previous COV for the SEP section, we are still concerned that the meaning, evaluation, and use of the Broader Impacts as a criterion for funding decisions is not well understood, or uniformly applied within and between programs.

Finally, while the Transformative and Interdisciplinary Research (FacTIR) criterion did not yet exist for the years under our charge (2005-2007), we are very concerned about how this will be understood and applied, both with NSF and by the PIs, reviewers, panels and POs. We strongly urge to the NSF to develop a strategy for making sure the definition is sensible, easily understood, and easily applied. We know there are efforts in this direction currently underway, but we would like to reinforce the importance of these activities.

C.4. Please provide comments on any other issues the COV feels are relevant.

None.

C.5. NSF would appreciate your comments on how to improve the COV review process, format and report template.

By and large we were impressed with the comprehensiveness of the data collected for the COV's benefit, and the openness and willingness of the POs and staff to provide data upon request.

We found a reporting problem in the scores for proposals and in the counting of the number of reviews. If a mail reviewer did not offer a numeric score, or elected to report a split number (e.g. very good/good), neither the review, nor the score was recorded, reducing the number of reviews reported and affecting in an unknown way the average scores reported. More importantly, it appears for some programs, panelists, or the subset of panelists assigned the proposal for review, were counted as reviewers, sometimes grossly inflating the reported number of reviewers. This is significant because in the data provided we noticed a disturbing strong negative correlation between the number of reviewers and the chance that a proposal was funded. In short, we wasted some of our valuable time discussing the spurious issues raised by this reporting error.

We also think that the metric of using the average for all metrics (proposal scores, award sizes and award duration) does not correctly indicate the central tendency because the distribution functions for these metrics are highly skewed. It would be more useful in the future to see the actual distribution of data, and to report, at least, the median as well as the mean.

The SEP program now supports several larger-scale initiatives which demand administrative time of the POs, and in some cases compete for resources from core SEP budgets. The COV needs to be provided with more explicit information about these programs including: 1) PO administrative duties, 2) review process for establishing the initiative (if formed within the review period), 3) review process of projects funded within the initiative, 4) source of funding (new or derived from core funding), 5) evidence of success of initiatives, and 6) discussion of planned future of the initiative.

Finally, we think it would be useful to have information of funding rates between first submittals and resubmissions.